G. IVCHENKO

YU. MEDVEDEV

A. CHISTYAKOV

PROBLEMS IN MATHEMATICAL STATISTICS

MIR PUBLISHERS MOSCOW

PROBLEMS IN MATHEMATICAL STATISTICS

Г. И. Ивченко, Ю. И. Медведев,
А. В. Чистяков

Сборник задач по математической статистике

«Высшая школа» Москва

G. I. Ivchenko, Yu. I. Medvedev, A. V. Chistyakov

PROBLEMS IN MATHEMATICAL STATISTICS

Mir Publishers Moscow

Translated from the Russian
by Elena Troshneva

First published 1991
Revised from the 1989 Russian edition

На английском языке

Printed in the Union of Soviet Socialist Republics

ISBN 5-03-001538-8
ISBN 5-06-000049-4

Contents

Preface

This problem book covers all the traditional topics in modern statistical theory and is designed for students at technical colleges and universities who have mathematical statistics as an obligatory course.

The problems are mostly analytical. The student is asked to prove the validity of an assertion or carry out an investigation. This will help him grasp the main aspects of mathematical statistics. Some of the problems are more difficult and can be used as individual assignments for course papers.

We have included problems on computer simulation of random variables in order to obtain the data for statistical interpretation. Any "theoretical" problem which contains a statistical algorithm for data analysis can be used (with the appropriate (practically infinite) choice of the model parameters) to formulate a "practical" problem. At the first stage the original data should be simulated using either published tables of random numbers or special computer programs. Then, by interpreting these "experimental" results according to the algorithm in question, the student can compare the theoretical hypothesis with the original parameters which are known as they were used when the sample was simulated.

All the problems differ in complexity. More difficult problems are marked with an asterisk and may require a significant effort on the part of the reader. Problems that cannot be reduced to standard algorithms are answered in detail or hints are given.

Each chapter contains the basic notions, assertions, and formulas from the respective theoretical section. The statistical tables at the end of the book will help the reader obtain numerical results. The list of distributions will help him choose problems on different aspects of the same model.

The Authors

THEORY AND PROBLEMS

CHAPTER 1

Principles of Statistical Description. Sampling Characteristics and Their Distributions

1.1. Problems in mathematical statistics are based on statistical data obtained by observations on a finite set of random variables $\mathbf{X} = (X_1, \ldots, X_n)$ which describe the outcome of an experiment. We say that the experiment consists of n trials, where the ith trial results in a random variable X_i, $i = 1, \ldots, n$. A set of observable random variables $\mathbf{X} = (X_1, \ldots, X_n)$ is called a *sample*, the values X_i, $i = 1, \ldots, n$, are called the *elements* (*units*) *of a sample*, and the number n is called the *sample size*. A set $\mathscr{X} = \{\mathbf{x} = (x_1, \ldots, x_n)\}$ of all possible realizations of the sample $\mathbf{X} = (X_1, \ldots, X_n)$ is called a *sample space*. When the true distribution of $\mathbf{X}$ (the distribution function $F_{\mathbf{X}}(x_1, \ldots, x_n) = \mathbf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n)$) is unknown (completely or partially) and only the class (family) of admissible distributions $\mathscr{F} = \{F(x_1, \ldots, x_n)\}$ which contains the distribution $F_{\mathbf{X}}$ of the sample $\mathbf{X}$ is specified, then we have a *statistical model* $(\mathscr{X}, \mathscr{F})$ (or simply *model* $\mathscr{F}$). Mathematical statistics reveals (within a given model $\mathscr{F}$) the properties of the true distribution $F_{\mathbf{X}}$ using the results of observations on the sample $\mathbf{X}$.

Some experiments consist of repeated independent observations on a random variable ξ (with the distribution $\mathscr{L}(\xi)$). Then the sample $\mathbf{X} = (X_1, \ldots, X_n)$ is a set of independent similarly distributed random variables, where $\mathscr{L}(X_i) = \mathscr{L}(\xi)$, $i = 1, \ldots, n$. To be concise, we say that $\mathbf{X} = (X_1, \ldots, X_n)$ is a *sample from the distribution* $\mathscr{L}(\xi)$. The statistical model for repeated independent observations is written as $\mathscr{F} = \{F_\xi\}$, i.e., we only indicate the class of admissible distribution functions of the original random variable ξ.

If $\mathscr{F} = \{F(x; \theta), \theta \in \Theta\}$, i.e., the admissible distribution functions are defined up to a parameter θ, then the model is said to be *parametric*, and the set Θ of the possible values of θ is called a *parametric set*.

We will only consider absolutely continuous or discrete models and use $f_\xi(x) = f(x)$ ($f(x; \theta)$ for parametric models) to denote the distribu-

tion density of the random variable ξ if the distribution F_ξ is absolutely continuous, and the probability $\mathbf{P}(\xi = x)$ if it is discrete.

In the case of a parametric model the distribution of probabilities on a sample space $\mathscr{X}$ which corresponds to the parameter θ is denoted $\mathbf{P}_\theta$. Similarly, $\mathbf{E}_\theta T(\mathbf{X})$, $\mathbf{D}_\theta T(\mathbf{X})$ are used to denote the moments of a given function $T(\mathbf{X})$ of the sample $\mathbf{X}$ when $F_{\mathbf{X}}(\mathbf{x}; \theta)$ is the distribution function of the sample.

1.2. Many problems in mathematical statistics concern sequences of random variables $\{\eta_n\}$ which converge to a limit η (a random variable or a constant) as $n \to \infty$. We will use two forms of convergence, i.e., convergence in probability ($\eta_n \xrightarrow{\mathbf{P}} \eta \Leftrightarrow \mathbf{P}(|\eta_n - \eta| > \varepsilon) \to 0$ $\forall \varepsilon > 0$) and convergence in distribution, or weak convergence ($\mathscr{L}(\eta_n) \to \mathscr{L}(\eta)$ or $\eta_n \xrightarrow{\mathscr{L}} \eta \Leftrightarrow F_{\eta_n}(x) \to F_\eta(x)$ $\forall x \in C(F_\eta)$, where $C(F)$ is the set of points of continuity of the function $F(x)$). Note that the $\mathbf{P}$-convergence implies the $\mathscr{L}$-convergence. The inference on the $\mathbf{P}$-convergence of various sampling characteristics often follows from the general assertion on the convergence of functions of random variables [7, p. 27], i.e., *if* $\eta_{ni} \xrightarrow{\mathbf{P}} c_i = \text{const}$, $i = 1, \ldots, r$, *and* $\varphi(x_1, \ldots, x_r)$ *is an arbitrary function continuous in the neighbourhood of the point* $(c_1, \ldots, c_r)$, *then* $\varphi(\eta_{n1}, \ldots, \eta_{nr}) \xrightarrow{\mathbf{P}} \varphi(c_1, \ldots, c_r)$.

1.3. If $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample from a distribution $\mathscr{L}(\xi)$, then $F_\xi(x) = F(x)$ is called a *theoretical distribution function*, and

$$F_n(x) = \frac{\mu_n(x)}{n} = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leqslant x) \tag{1.1}$$

is an *empirical distribution function* (here $\mu_n(x)$ is the number of elements in a sample, which satisfy the condition $X_j \leqslant x$, and $I(A)$ is the indicator of the event A).

By the Bernoulli theorem, $F_n(x) \xrightarrow{\mathbf{P}} F(x)$ $\forall x$ as $n \to \infty$, i.e., for large n the value of $F_n(x)$ can be an estimate for $F(x)$. The Glivenko and Kolmogorov theorems on the asymptotic properties of $F_n(x)$ for large n [7, p. 22] prove that the empirical distribution function can be an estimator for the theoretical distribution function.

If a random variable ξ is discrete and assumes the values a_1, $a_2, \ldots$, then the distribution law for ξ may be conveniently represented by the frequencies h_r/n, where h_r is the number of units in a sample, which are equal to a_r. Then $h_r/n \xrightarrow{\mathbf{P}} \mathbf{P}(\xi = a_r)$, $r = 1, 2, \ldots$, as $n \to \infty$.

If the values of ξ have the density $f_\xi(x) = f(x)$, we may investigate the frequencies h_k/n of the events $\{\xi \in \Delta_k\}$, where $\{\Delta_k\}$ is a system of nonintersecting intervals into which the region of the possible ξ-

values is divided. Then

$$\frac{h_k}{n} \xrightarrow{P} \mathbf{P}(\xi \in \Delta_k) = \int_{\Delta_k} f(x)\,dx$$

as $n \to \infty$, and, if Δ_k are small, we may use the frequencies h_k/n to construct a *histogram* and a *frequency polygon* which are close to the graph of the function $f(x)$ [7, p. 23] and give an approximate form of the distribution of ξ.

Every theoretical characteristic $g = \int g(x)\,dF(x)$ corresponds to its *statistical analogue (copy)*

$$G = G(\mathbf{X}) = \int g(x)\,dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

which is called the *empirical* or *sampling characteristic*. Specifically, sampling moments are statistical analogues for theoretical moments. The quantity

$$A_{nk} = A_{nk}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

is a *sampling moment of kth order*. At $k = 1$ the quantity A_{n1} is called a *sample mean* and is denoted $\overline{X}$, viz.,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The quantity

$$M_{nk} = M_{nk}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k$$

is called the *central sampling moment of kth order*. At $k = 2$ the quantity M_{n2} is called the *sample variance* and is denoted $S^2 = S^2(\mathbf{X})$, viz.,

$$S^2(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2.$$

The notation $S'^2 = \dfrac{n}{n-1} S^2$ may also be used. The absolute sam-

pling moments, sampling semi-invariants, etc., are introduced in a similar way.

Sample quantiles are another example of sampling characteristics. A p-quantile for any distribution function $F(x)$ is defined as $\zeta_p = \inf\{x: F(x) \geqslant p\}$, $0 < p < 1$, and a *sample p-quantile* $Z_{n,p}$ is a p-quantile of the empirical distribution function $F_n(x)$. If the units of the sample $\mathbf{X} = (X_1, \ldots, X_n)$ are arranged in increasing order of magnitude, we get a new sequence of random variables

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$$

which is called an *ordered series of a sample*. Here $X_{(k)}$ is the *kth-order statistic*, $k = 1, \ldots, n$, and $X_{(1)}$ and $X_{(n)}$ are the *extrema of the sample*. Then we can express $Z_{n,p}$ through order statistics

$$Z_{n,p} = \begin{cases} X_{([np]+1)} & \text{for non-integer } np, \\ X_{(np)} & \text{for integer } np. \end{cases}$$

Specifically, $Z_{n,1/2}$ is a *sample median*.

Any sampling characteristic which is a continuous function of a finite number of the values A_{nk} (in particular, the sampling moments and central sampling moments M_{nk}) converges in probability to the respective theoretical characteristic as $n \to \infty$ and can be an *estimator* for the latter when the number n of observations is sufficiently large. Similarly, $Z_{n,p} \xrightarrow{\mathrm{P}} \zeta_p$ if only the distribution $\mathscr{L}(\xi)$ has a smooth density.

1.4. The sampling theory studies various properties of the distribution of sampling characteristics in exact and asymptotic (for large sample sizes) forms. When investigating the asymptotic behaviour (as $n \to \infty$) of distributions, the limit theorems of probability theory (specifically, the law of large numbers and the Central Limit Theorem) are frequently used. We take their simplest forms from [2].

The law of large numbers. *If the random variables $\eta_1, \eta_2, \ldots, \eta_n$ are independent, similarly distributed, and their expected values are* $\mathbf{E}\eta_i = a$, *then as* $n \to \infty$

$$\frac{1}{n}(\eta_1 + \ldots + \eta_n) \xrightarrow{\mathrm{P}} a.$$

The Central Limit Theorem. *If in addition to the above conditions there exists* $\mathbf{D}\eta_i = \sigma^2 > 0$, *then as* $n \to \infty$

$$\mathscr{L}((\eta_1 + \ldots + \eta_n - na)/(\sqrt{n}\sigma)) \to \mathscr{N}(0, 1).$$

A multi-dimensional version of the Central Limit Theorem has the form: *let the r-dimensional random vectors* $\eta_n = (\eta_{n1}, \ldots, \eta_{nr})$, $n = 1$,

2, ..., *be independent, similarly distributed, and have finite moments*

$$a_i = \mathbf{E}\eta_{1i}, \quad b_{ij} = \operatorname{cov}(\eta_{1i}, \eta_{1j}), \ i, j = 1, \ldots, r.$$

Then as $n \to \infty$

$$\mathscr{L}(\zeta_{n1}, \ldots, \zeta_{nr}) \to \mathscr{N}(0, \ \mathbf{B} = \|b_{ij}\|_1^r),$$

where

$$\zeta_{ni} = (\eta_{1i} + \ldots + \eta_{ni} - na_i)/\sqrt{n}, \quad i = 1, \ldots, r.$$

(The definition of a multi-dimensional (multivariate) normal distribution see in Sec. 1.6.)

Let us formulate some assertions on the convergence of the functions of random variables, which we will need to solve problems.

1°. *If* $\eta_n \xrightarrow{\mathbf{P}} \eta$ *and the function* φ *is continuous, then* $\varphi(\eta_n) \xrightarrow{\mathbf{P}, \mathscr{L}} \varphi(\eta)$.

2°. *Let* $\{\eta_n, \zeta_n\}$, $n = 1, 2, \ldots$, *be a sequence of pairs of random variables. Then*

(a) $\eta_n - \zeta_n \xrightarrow{\mathbf{P}} 0$, $\zeta_n \xrightarrow{\mathbf{P}, \mathscr{L}} \zeta \Rightarrow \eta_n \xrightarrow{\mathbf{P}, \mathscr{L}} \zeta$;

(b) $\mathscr{L}(\eta_n) \to \mathscr{L}(\eta)$, $\zeta_n \xrightarrow{\mathbf{P}} 0 \Rightarrow \eta_n\zeta_n \xrightarrow{\mathbf{P}} 0$;

(c) $\mathscr{L}(\eta_n) \to \mathscr{L}(\eta)$, $\zeta_n \xrightarrow{\mathbf{P}} c = \text{const} \Rightarrow \mathscr{L}(\eta_n + \zeta_n) \to \mathscr{L}(\eta + c)$, $\mathscr{L}(\eta_n\zeta_n) \to \mathscr{L}(c\eta)$, $\mathscr{L}(\eta_n/\zeta_n) \to \mathscr{L}(\eta/c)$ *for* $c \neq 0$;

(d) $\eta_n - \zeta_n \xrightarrow{\mathbf{P}} 0$, $\mathscr{L}(\zeta_n) \to \mathscr{L}(\zeta)$, *the function* φ *is continuous* $\Rightarrow \varphi(\eta_n) - \varphi(\zeta_n) \xrightarrow{\mathbf{P}} 0$.

3°. *Let* $T_n = T_n(\mathbf{X})$, $\mathbf{X} = (X_1, \ldots, X_n)$, *be the estimator of a scalar parameter* θ *in the model* $\mathscr{F} = \{F(x; \theta), \ \theta \in \Theta\}$ *such that* $\mathscr{L}_\theta(\sqrt{n}(T_n - \theta)) \to \mathscr{N}(0, \sigma^2(\theta))$ *as* $n \to \infty$ *and for all* $\theta \in \Theta$. *Suppose also that the function* φ *is differentiable and* $\varphi' \neq 0$. *Then*

$$\mathscr{L}_\theta(\sqrt{n}[\varphi(T_n) - \varphi(\theta)]) \to \mathscr{N}(0, \ [\varphi'(\theta)]^2\sigma^2(\theta)).$$

Besides, if the functions φ' *and* σ *are continuous, then*

$$\mathscr{L}_\theta\left(\sqrt{n}\,\frac{\varphi(T_n) - \varphi(\theta)}{\varphi'(T_n)\sigma(T_n)}\right) \to \mathscr{N}(0, \ 1).$$

The generalization of 3° to the case of a vector parameter $\theta = (\theta_1, \ldots, \theta_r)$ has the following form.

4°. *Let* $\mathbf{T}_n = (T_{n1}, \ldots, T_{nr})$ *be an estimator of the parameter* θ *satisfying the condition* $\mathscr{L}_\theta(\sqrt{n}(\mathbf{T}_n - \theta)) \to \mathscr{N}(0, \Sigma(\theta))$ *as* $n \to \infty$ *for all* $\theta \in \Theta$. *Then for any differentiable function* φ *of* r *variables we have*

$$\mathscr{L}_\theta(\sqrt{n}(\varphi(\mathbf{T}_n) - \varphi(\theta))) \to \mathscr{N}(0, \ v^2(\theta))$$

under the condition that $v(\theta) \neq 0$, *where* $v^2(\theta) = \mathbf{b}'(\theta)\Sigma(\theta)\mathbf{b}(\theta)$, $\mathbf{b}(\theta) = \left(\dfrac{\partial\varphi}{\partial\theta_1}, \ldots, \dfrac{\partial\varphi}{\partial\theta_r}\right)$. *Moreover, if the function* φ *is continuously*

differentiable and all the elements of the matrix of the second moments $\Sigma(\theta)$ *are continuous in* θ, *then*

$$\mathscr{L}_\theta(\sqrt{n}[\varphi(\mathbf{T}_n) - \varphi(\theta)]/v(\mathbf{T}_n)) \to \mathscr{N}(0, 1).$$

By the Central Limit Theorem the sampling moment A_{nk} is asymptotically normal and its parameters are $\alpha_k = \mathbf{E}\xi^k$ and $\frac{1}{n}\mathbf{D}\xi^k = (\alpha_{2k} - \alpha_k^2)/n$, which may be briefly written as $\mathscr{L}(A_{nk}) \sim \mathscr{N}(\alpha_k, (\alpha_{2k} - \alpha_k^2)/n)$. The joint distribution of any finite number of sampling moments A_{nk} is also asymptotically normal, as well as (under some additional conditions) the distribution of any differentiable function of a finite number of moments A_{nk}. Specifically, central sampling moments M_{nk} are also asymptotically normal.

We use direct analysis of exact distributions of order statistics $X_{(k)}$ to investigate the asymptotic behaviour of $X_{(k)}$ as $n \to \infty$. For the distributions $\mathscr{L}(\xi)$ with smooth densities the *mid terms* of the ordered series (i.e., when the number $k = k(n)$ satisfies the condition $k/n \to p$, $0 < p < 1$) are asymptotically normal, while for the *extreme* order statistics (i.e., for $X_{(r)}$, $X_{(n-s+1)}$ at fixed $r, s \geqslant 1$) the class of limiting distributions only consists of three types of distributions, which are not normal [7, p. 35].

1.5. Some formulas of probability theory, which are used to obtain an explicit form of a distribution when transforming the random variables, are appropriate here. Let the vector $\mathbf{X} = (X_1, \ldots, X_k)$ have an absolutely continuous distribution with the density $f(\mathbf{x})$, $\mathbf{x} = (x_1, \ldots, x_k) \in S \subseteq R^k$ and let $\mathbf{h} = (h_1, \ldots, h_k)$: $S \to R^k$ be an arbitrary, one-to-one, and smooth (i.e., all its partial derivatives $\partial h_i(\mathbf{x})/\partial x_j$ are continuous) transformation whose Jacobian

$$J(\mathbf{x}) = \det \begin{Vmatrix} \dfrac{\partial h_1(\mathbf{x})}{\partial x_1} & \cdots & \dfrac{\partial h_k(\mathbf{x})}{\partial x_1} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \dfrac{\partial h_1(\mathbf{x})}{\partial x_k} & \cdots & \dfrac{\partial h_k(\mathbf{x})}{\partial x_k} \end{Vmatrix}$$

does not vanish on S. Then the distribution density of the random vector $\mathbf{Y} = \mathbf{h}(\mathbf{X}) = (h_1(\mathbf{X}), \ldots, h_k(\mathbf{X}))$ has the form

$$\varphi(\mathbf{y}) = f(\mathbf{h}^{-1}(\mathbf{y}))/|J(\mathbf{h}^{-1}(\mathbf{y}))|, \quad \mathbf{y} = (y_1, \ldots, y_k) \in \mathbf{h}(S), \qquad (1.2)$$

where $\mathbf{h}^{-1}$ is a transformation inverse to $\mathbf{h}$, i.e., $\mathbf{h}^{-1}(\mathbf{h}(\mathbf{x})) \equiv \mathbf{x}$. Two special cases are frequently encountered. If $k = 1$, then we have to

transform the random variable $Y = h(X)$, where $h(x)$ is a one-to-one smooth function with a non-vanishing derivative. Then the distribution density of Y has the form

$$\varphi(y) = f(h^{-1}(y))/|h'(h^{-1}(y))|. \tag{1.3}$$

If we have a linear transformation $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, $\det \mathbf{A} \equiv \mathbf{a} \neq 0$, then the distribution density of $\mathbf{Y}$ is $\varphi(\mathbf{y}) = f(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}))/|\mathbf{a}|$.

Some statistical problems deal with the ratio $\zeta = \xi/\eta$ of two independent random variables whose distribution densities f_ξ and f_η are known. The distribution density of ζ can be found from the formula

$$f_\zeta(y) = \int_{-\infty}^{\infty} f_\xi(xy) f_\eta(x)|x|\, dx. \tag{1.4}$$

1.6. We will need some frequently applied distributions $\mathscr{L}(\xi)$ and their properties.

(1) The *normal distribution* $\mathcal{N}(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$, has the density $\dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$. Here $\mu = \mathbf{E}\xi$, $\sigma^2 = \mathbf{D}\xi$, and the central moments $\mu_k = \mathbf{E}(\xi - \mu)^k$ are $\mu_{2r+1} = 0$, $\mu_{2r} = \dfrac{(2r)!}{r!2^r}\sigma^{2r} = 1 \times 3 \ldots (2r-1)\sigma^{2r}$, respectively. The distribution $\mathcal{N}(0, 1)$ is called a *standard normal distribution*; its distribution function is $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-t^2/2}\, dt$, the equation $\Phi(u_p) = p$, $p \in (0, 1)$, uniquely defines its p-quantile u_p with $u_{1-p} = -u_p$. The notation $c_\gamma = u_{(1+\gamma)/2}$ can also be found in the literature. The random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ has a k-variate normal distribution $\mathcal{N}(\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k), \boldsymbol{\Sigma} = |\sigma_{ij}|_1^k)$ if its characteristic function is of the form*

$$\mathbf{E}e^{i\mathbf{t}'\boldsymbol{\xi}} = \exp\left\{i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right\},$$

$$\mathbf{t} = (t_1, \ldots, t_k).$$

Here

$$\mathbf{E}(\boldsymbol{\xi}) \equiv (\mathbf{E}\xi_1, \ldots, \mathbf{E}\xi_k) = \boldsymbol{\mu},$$

$$\mathbf{D}(\boldsymbol{\xi}) \equiv \mathbf{E}(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})' \equiv |\mathrm{cov}\,(\xi_i, \xi_j)|_1^k = |\sigma_{ij}|_1^k = \boldsymbol{\Sigma}.$$

* In matrix operations vectors are treated as column-vectors and ′ stands for a transposition.

If $|\Sigma| \neq 0$, the distribution $\mathcal{N}(\mu, \Sigma)$ is *non-degenerate* and has the density

$$f(\mathbf{x}) = (2\pi)^{-k/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\},$$
$$\mathbf{x} = (x_1, \ldots, x_k) \in R^k.$$

The normal distribution has an important property that under a linear transformation $\eta = \mathbf{A}\xi$ ($\mathbf{A}$ is a given matrix) we obtain a normal random vector and $\mathcal{L}(\eta) = \mathcal{N}(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}')$. Specifically, if $\eta = \mathbf{U}'\xi$, where $\mathbf{U}$ is an orthogonal matrix which reduces Σ to a diagonal form $\mathbf{U}'\Sigma\mathbf{U} = \mathbf{D} = \begin{Vmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_k\end{Vmatrix}$ (λ_j, $j = 1, \ldots, k$, are the eigenvalues of Σ), then $\mathcal{L}(\eta) = \mathcal{N}(\mathbf{U}'\mu, \mathbf{D})$, i.e., the components of the vector η are non-correlated and therefore independent. Putting $\mathbf{Z} = \mathbf{D}^{-1/2}\mathbf{U}'(\xi - \mu)$ (if all the $\lambda_j > 0$), we get $\mathcal{L}(\mathbf{Z}) = \mathcal{N}(\mathbf{0}, \mathbf{E}_k)$, where $\mathbf{E}_k$ is an identity matrix of dimension k. Thus, we can always find a linear transformation to turn a non-degenerate normal vector into a vector with independent standard normal components.

When applying samples from normal distributions, we need the following important assertions [7, pp. 38-40].

1°. *If* $\mathbf{X} = (X_1, \ldots, X_n)$ *is a sample from the distribution* $\mathcal{N}(\mu, \sigma^2)$ *and* $\mathbf{t} = \mathbf{BX}$, $Q_i = \mathbf{X}'\mathbf{A}_i\mathbf{X}$, $i = 1, 2$, *are, respectively, linear and quadratic functions of* $\mathbf{X}$, *then it is sufficient that* $\mathbf{BA}_i = \mathbf{0}$ *for* $\mathbf{t}$ *and* Q_i *to be independent, and* $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1 = \mathbf{0}$ *for* Q_1 *and* Q_2 *to be independent.*

2°. *Let* $\mu = 0$, $\sigma^2 = 1$, *and* $\mathbf{A}_1^2 = \mathbf{A}_1$ (*the matrix* $\mathbf{A}_1$ *is idempotent*). *Then* $\mathcal{L}(Q_1) = \chi^2(r)$, *where* $r = \operatorname{rank} \mathbf{A}_1 = \operatorname{tr} \mathbf{A}_1$ *is the trace of the matrix* $\mathbf{A}_1$.

3°. **Fisher's theorem.** *The sample mean* $\overline{X}$ *and variance* S^2 *are independent and* $\mathcal{L}(\sqrt{n}(\overline{X} - \mu)/\sigma) = \mathcal{N}(0, 1)$, $\mathcal{L}(nS^2/\sigma^2) = \chi^2(n-1)$. (The definition of the χ^2-distribution will be given below.)

(2) The *gamma distribution* $\Gamma(a, \lambda)$, $a, \lambda > 0$, is defined by the density $\dfrac{x^{\lambda-1}e^{-x/a}}{\Gamma(\lambda)a^\lambda}$, $x > 0$ $\left(\text{here } \Gamma(\lambda) = \int_0^\infty t^{\lambda-1}e^{-t}\,dt,\ \lambda > 0, \text{ is a } \textit{gamma function}\right)$, and its moments are $\mathbf{E}\xi^b = a^b\Gamma(\lambda + b)/\Gamma(\lambda)$, $b > -\lambda$. In particular, $\mathbf{E}\xi = a\lambda$, $\mathbf{D}\xi = a^2\lambda$.

The special case $\Gamma(a, 1)$ is called an *exponential distribution.* Another special case is $\Gamma(2, n/2)$. It is called the *chi-square distribution with n degrees of freedom* and is denoted $\chi^2(n)$. Here $\chi^2(n) =$

$\mathscr{L}(\xi_1^2 + \ldots + \xi_n^2)$, the terms are independent, and $\mathscr{L}(\xi_i) = \mathscr{N}(0, 1)$, $i = 1, \ldots, n$. We use $\chi^2_{p,n}$ to denote the p-quantiles of the distribution $\chi^2(n)$.

(3) In the general case, *Weibull's distribution* $W(a, \alpha, b)$ depends on three parameters, i.e., the *location (position) parameter* $a \in R^1$, the *shape parameter* $\alpha > 0$, and the *scale parameter* $b > 0$, and is defined by the distribution function

$$F_\xi(x) = 1 - \exp\left\{-\left(\frac{x-a}{b}\right)^\alpha\right\}, \quad x \geqslant a.$$

Here

$$\mathbf{E}\xi = a + b\Gamma\left(1 + \frac{1}{\alpha}\right),$$

$$\mathbf{D}\xi = b^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right].$$

The special case $W(a, 1, b)$ is known as a *two-parameter exponential distribution*, and the case $W(a, 2, b)$ is known as *Rayleigh's distribution.*

(4) The *beta distribution* $B(a, b)$, $a, b > 0$, is defined by the density $x^{a-1}(1-x)^{b-1}/\mathrm{B}(a, b)$, $0 \leqslant x \leqslant 1$, where $\mathrm{B}(a, b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is a *beta function.* Here

$$\mathbf{E}\xi = \frac{a}{a+b}, \quad \mathbf{D}\xi = \frac{ab}{(a+b)^2(a+b+1)}.$$

(5) The *uniform distribution* $R(a, b)$, $-\infty < a < b < \infty$, has a constant density $f(x) = \dfrac{1}{b-a}$, $a \leqslant x \leqslant b$. Here

$$\mathscr{L}\left(\frac{\xi - a}{b - a}\right) = R(0, 1) = B(1, 1),$$

$$\mathbf{E}\xi = \frac{a+b}{2}, \quad \mathbf{D}\xi = \frac{(b-a)^2}{12}.$$

(6) *Cauchy's distribution* $C(a)$, $-\infty < a < \infty$, is defined by the density $\dfrac{1}{\pi}\dfrac{1}{1+(x-a)^2}$, $-\infty < x < \infty$. This distribution has no mo-

ments (including the mathematical expectation), and the constant a coincides with the median $\zeta_{1/2}$. Cauchy's distribution has an important property that *if the random variables* $\xi_1, \ldots, \xi_n$ *are independent and* $\mathscr{L}(\xi_i) = C(a_i)$, $i = 1, \ldots, n$, *then* $\mathscr{L}(\bar{\xi}) = C(\bar{a})$, where the bar denotes an *arithmetic mean.*

(7) *Student's distribution* $S(n) \equiv \mathscr{L}(t_n \equiv \eta/\sqrt{\chi_n^2/n})$ *with n degrees of freedom,* where η and χ_n^2 are independent random variables and $\mathscr{L}(\eta) = \mathscr{N}(0, 1)$, $\mathscr{L}(\chi_n^2) = \chi^2(n)$, has a density of the form

$$s_n(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}, \quad -\infty < x < \infty.$$

We use $t_{p,n}$ to denote its p-quantiles.

(8) *Snedecor's distribution* $S(n_1, n_2) \equiv \mathscr{L}\left(F_{n_1,n_2} \equiv \frac{\chi_{n_1}^2}{n_1} \div \frac{\chi_{n_2}^2}{n_2}\right)$ *with* n_1 *and* n_2 *degrees of freedom,* where $\chi_{n_1}^2$ and $\chi_{n_2}^2$ are independent random variables, and $\mathscr{L}(\chi_{n_i}^2) = \chi^2(n_i)$, $i = 1, 2$, has a density of the form

$$f_{n_1,n_2}(x) = \left(\frac{n_1}{n_2}\right)^{n_1/2} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \frac{x^{n_1/2-1}}{\left(1 + \frac{n_1}{n_2}x\right)^{(n_1+n_2)/2}}, \quad x > 0.$$

We use F_{p,n_1,n_2} to denote its p-quantile, where $F_{1-p,n_1,n_2} = 1/F_{p,n_2,n_1}$.

(9) The *binomial distribution* $Bi(n, p)$ is a distribution of the number of successes in n independent trials with two outcomes (success-failure) and a constant probability of success $p \in (0, 1)$ (the Bernoulli trials). Here

$$\mathbf{P}(\xi = k) = C_n^k p^k q^{n-k}, \ k = 0, 1, \ldots, n, \ q = 1 - p,$$
$$\mathbf{E}\xi = np, \quad \mathbf{D}\xi = npq.$$

For $n = 1$ we have *Bernoulli's distribution* $Bi(1, p)$.

(10) The *polynomial distribution* $M(n; p_1, \ldots, p_N)$, $p_1 + \ldots + p_N = 1$, is a distribution of a random vector $\nu = (\nu_1, \ldots, \nu_N)$ with non-negative integer-valued components satisfying the condition

$\nu_1 + \ldots + \nu_N = n$ which has the form

$$\mathbf{P}(\nu = \mathbf{h}) = \frac{n!}{h_1! \ldots h_N!} p_1^{h_1} \ldots p_N^{h_N},$$

$$\mathbf{h} = (h_1, \ldots, h_N), \quad h_1 + \ldots + h_N = n.$$

Here

$$\mathbf{E}\nu_i = np_i, \quad \operatorname{cov}(\nu_i, \nu_j) = \begin{cases} np_i(1 - p_i) & \text{for} \quad i = j, \\ -np_ip_j & \text{for} \quad i \neq j. \end{cases}$$

If we carry out n independent trials with N possible outcomes whose probabilities do not vary and are equal to $p_1, \ldots, p_N$, respectively, then, by using ν_i to denote the number of the realizations of the ith outcome, $i = 1, \ldots, N$, we will obtain $\mathcal{L}(\nu) = M(n; p_1, \ldots, p_N)$. If $N = 2$, we have $M(n; p, 1 - p) = Bi(n, p)$, i.e., the polynomial distribution is reduced to the binomial one.

(11) *Poisson's distribution* $\Pi(\lambda)$, $\lambda > 0$, is defined by the probabilities

$$\mathbf{P}(\xi = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

Here $\lambda = \mathbf{E}\xi = \mathbf{D}\xi$ and, in general, $\mathbf{E}(\xi)_j = \lambda^j$, where $(a)_j = a(a - 1)\ldots(a - j + 1)$, $j \geqslant 1$, $(a)_0 = 1$.

(12) The *negative binomial distribution* $\overline{Bi}(r, p)$, $p \in (0, 1)$, $r = 1, 2, \ldots$, is defined by the probabilities

$$\mathbf{P}(\xi = k) = C_{r+k-1}^k p^k q^r, \quad k = 0, 1, 2, \ldots, q = 1 - p.$$

This is a distribution of the number of successes before the rth failure in an infinite sequence of Bernoulli trials. Here

$$\mathbf{E}\xi = rp/q, \quad \mathbf{D}\xi = rp/q^2.$$

In the special case of $r = 1$, the $\overline{Bi}(1, p)$ distribution is called a *geometric disribution.*

(13) The *hypergeometric distribution* $H(r, N, n)$ is defined by the probabilities

$$\mathbf{P}(\xi = k) = C_r^k C_{N-r}^{n-k} / C_N^n,$$

$$\max(0, n + r - N) \leqslant k \leqslant \min(n, r).$$

If an urn contains N balls r of which are red and $N - r$ are black, and we withdraw from it without replacement a random sample of size n, then the random variable ξ (the number of red balls in the

sample) has a hypergeometric distribution. Here

$$\mathbf{E}\xi = \frac{nr}{N}, \quad \mathbf{D}\xi = \frac{nr}{N}\left(1 - \frac{r}{N}\right)\frac{N-n}{N-1}$$

and, in general, $\mathbf{E}(\xi)_j = \dfrac{(n)_j(r)_j}{(N)_j}$.

Other properties of these distributions are considered in Problems 1.39-55.

If a statistical model $\mathscr{F} = \{F_\xi\}$ is defined by a standard distribution with unknown parameters θ (if there are a few parameters, only some of them may be unknown), the model preserves the name of the distribution. For example, the model $\mathscr{N}(\theta, \sigma^2)$ is said to be normal with an unknown mean, the model $\mathscr{N}(\mu, \theta^2)$ is normal with only variance as the unknown parameter, the model $\mathscr{N}(\theta_1, \theta_2^2)$ is a general normal model with two unknown parameters, the model $\Pi(\theta)$ is Poisson's model.

1.7. *Statistical simulation* using a sequence of pseudo-random numbers helps to illustrate the efficiency of various statistical procedures. *Pseudo-random numbers* are sequences of numbers obtained by a certain algorithm and having the properties of a sequence of random numbers. The methods of obtaining pseudo-random numbers can be found in [4, 9].

A realization of a sequence of arbitrarily distributed independent random numbers is commonly obtained from a realization of a sequence of independent random numbers uniformly distributed on a segment [0, 1].

A realization of the uniformly distributed random numbers

$$U_0,\ U_1,\ U_2,\ \ldots \tag{1.5}$$

is frequently obtained by the linear congruent method [9]

$$U_n = z_n/m, \tag{1.6}$$

where z_n is a sequence defined by the recurrence relation

$$z_{n+1} = az_n + c \pmod{m},$$

where z_0 is the initial value, and a, c, and m are positive integers.

Strictly speaking, the sequence (1.5) defined by (1.6) cannot be treated as a realization of an independent sequence of uniformly distributed numbers, because it is either periodic or periodic with a lead sequence. The length of the period T is less than m, because the number of different values of z_n, $n = 0, 1, 2, \ldots$, does not exceed m. It is obvious that the sequences which exceed the terms before the peri-

od plus the length of the period should not be used. Nevertheless, the sequence (1.5) can have the greatest possible period m when the constants a, c, m, and z_0 are chosen properly.

The following theorem defines the conditions for the period of a sequence to be maximal [9].

Theorem. *The length of the period of a linear congruent sequence* (1.5) *is equal to* m *if and only if*

(1) c *and* m *are relatively prime numbers;*

(2) $b = a - 1$ *is divisible by* p *for any prime* p *which is a divisor of* m;

(3) b *is divisible by* 4 *if* m *is divisible by* 4.

The presence of a complete period does not always ensure good properties of pseudo-random numbers. Even the commonly used generators have essential drawbacks. Various statistical tests [6] help to verify the "quality" of the sequences generated. It is usually enough to check whether the s-chains ($s = 1, 2, \ldots$) of the sequence (1.5) are uniformly distributed, and then use this sequence to solve simulation problems.

Let us simulate $n = 100$ uniformly distributed numbers $X_1, X_2, \ldots, X_n$ and list the results in Table 1.1.

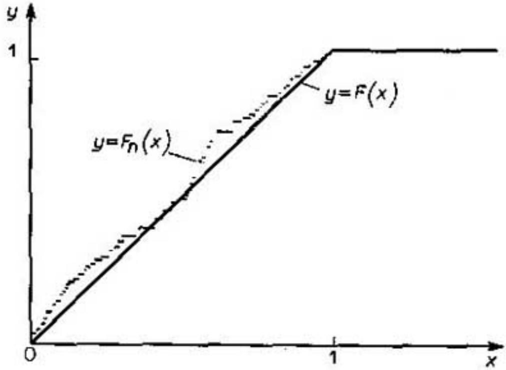
Table 1.1

0.168	0.273	0.878	0.983	0.588	0.693	0.298	0.403	0.008	0.113
718	823	428	533	138	243	848	953	558	663
549	754	459	664	369	574	279	484	189	394
099	304	009	214	919	124	829	034	739	944
550	855	660	965	770	075	880	185	990	295
100	405	210	515	320	625	430	735	540	845
571	976	881	286	191	596	501	906	811	216
121	526	431	836	741	146	051	456	361	766
012	517	522	027	032	537	542	047	052	557
562	067	072	577	562	087	092	597	602	107

Figure 1 shows the empirical distribution function $F_n(x)$ constructed from these data.

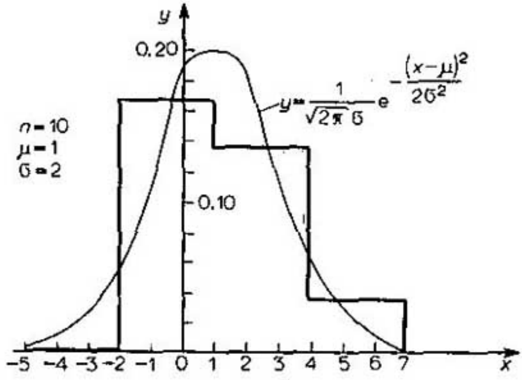
We now obtain n normally distributed random numbers $X_1, X_2, \ldots, X_n$ with the parameters $\mu = \mathbf{E}X_i$, $\sigma^2 = \mathbf{D}X_i$. The histrograms for $\mu = 1$, $\sigma^2 = 4$, and $n = 10, 100, 1000$ are plotted in Figs. 2-4.

We divide the xy-axis into intervals of length h, where $h = 3, 1.5, 0.75$ for $n = 10, 100, 1000$, respectively. The boundary point of the intervals is $x = 1$ for any n.



y
1
y=F(x)
y=F_n(x)
0
1
x

Fig. 1



y
0.20
0.10
n=10
μ=1
σ=2
y = 1/(√(2π) σ) e^(−(x−μ)²/(2σ²))
-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7
x

Fig. 2

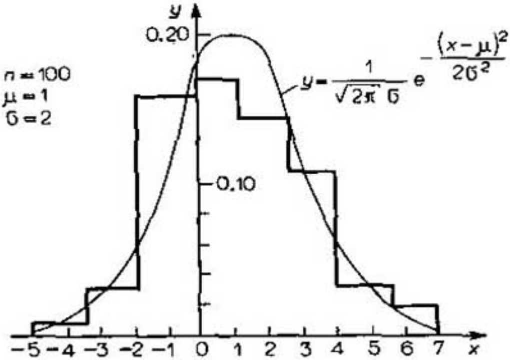

y
0.20
0.10
n = 100
μ = 1
σ = 2
y = 1/(√(2π) σ) e^(−(x−μ)²/(2σ²))
−5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 x

Fig. 3

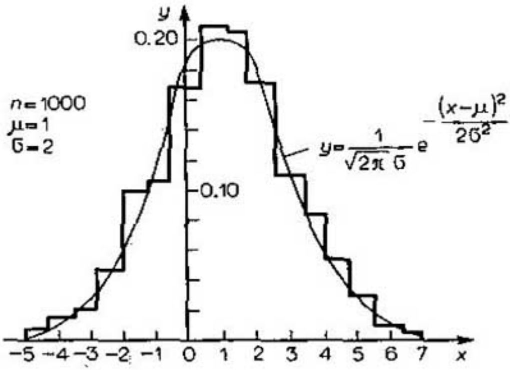

y
0.20
0.10
n = 1000
μ = 1
σ = 2
y = 1/(√(2π) σ) e^(−(x−μ)²/(2σ²))
−5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 x

Fig. 4

Table 1.2 gives the estimates

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

for the parameters $\mu = 1$ and $\sigma^2 = 4$.

Table 1.2

n	10	100	1000
$\overline{X}$	0.676	1.016	0.988
S'^2	3.901	4.315	4.306

Problems

1.1. Suggest a method to simulate a sequence of Bernoulli trials X_1, X_2, ..., X_n, ..., where $\mathbf{P}(X_n = 1) = 1 - \mathbf{P}(X_n = 0) = p$.

Hint. Use a sequence of pseudo-random numbers uniformly distributed on the segment $[0, 1]$.

1.2. Simulate a sequence of Bernoulli trials as in Problem 1.1, where $p = 0.4$ and $n = 1000$. Calculate the frequencies μ_k/k, where $\mu_k = X_1 + \ldots + X_k$, for $k = 100, 200, \ldots, 900, 1000$. Construct a graph in the xy-plane by connecting the neighbouring points $(k, \mu_k/k)$, $k = 100, 200, \ldots, 1000$, by straight lines.

1.3. Find a way to simulate independent trials in a polynomial scheme with the outcomes $1, 2, \ldots, N$ whose probabilities are p_1, p_2, ..., p_N, respectively.

1.4. Find a way to simulate a discrete-time symmetric wandering through integer points on a straight line with origin at the point 0 (the probabilities of transitions to neighbouring points in a single step are taken to be the same).

1.5. Let a random variable ξ be uniformly distributed on the interval $[0, 1]$, and let $F(x)$ be a continuous distribution function. Find the distribution function of the random variable $\eta = F^{-1}(\xi)$, where $x = F^{-1}(y)$ is a function inverse to $y = F(x)$.

1.6. Suggest a simulation technique for a random sequence X_1, X_2, ..., X_n, ..., where $\mathbf{P}(X_n \leqslant t) = 1 - e^{-t/a}$, $t \geqslant 0$ ($a > 0$ is a constant).

Hint. Use the previous problem.

1.7. Simulate independent and exponentially distributed quantities $X_1, X_2, \ldots, X_n$ with $a = 1$ and $n = 100$. Construct an empirical distribution function and a histogram. Calculate the first and second sampling moments A_{n1} and A_{n2}.

| *Hint.* Use the previous problem.

1.8. Suggest a simulation technique for an Erlang random sequence $\{X_j\}$ with the parameters (a, m) (i.e., $\mathcal{L}(X_j) = \Gamma(a, m)$, $j = 1, 2, \ldots$).

1.9. Using the Central Limit Theorem, find a way to simulate approximately normally distributed random numbers X_n, $n = 1, 2, \ldots$.

1.10. Let $X_{N_1}, \ldots, X_{N_n}$ be the realization of a sequence of approximately normally distributed numbers each of which is obtained by summing up N uniformly distributed terms (see the previous problem). Get three realizations (for $N = 2, 4, 12$) of the samples with $n = 100$, $a = 0$, and $\sigma^2 = 1$. Construct the empirical distribution functions and histograms for each sample. Find the estimates for a and σ^2.

1.11. Using the samples from the previous problem, calculate the third and fourth central sampling moments and compare them with the true values of the theoretical moments.

1.12. Suggest a simulation technique for a sample from a binomial distribution $Bi(k, p)$.

1.13. Let ν_n be the number of successes in n Bernoulli trials with the probability of success $p \in (0, 1)$. Under the condition that n is large calculate the boundary δ_γ such that the event $\left|\frac{\nu_n}{n} - p\right| \leqslant \delta_\gamma$ has the probability $\simeq \gamma$. Check whether the results of the following (De Buffon's) experiment lie within these boundaries for $\gamma = 0.98$, viz., heads appeared $h = 2048$ times at $n = 4040$ tossings of a coin.

| *Hint.* Apply the De Moivre-Laplace theorem and consider the coin to be symmetric.

1.14. Using the approach of the previous problem, check whether the following data correspond to the theory, viz., among the $n = 10\,000$ randomly placed numbers $0, 1, \ldots, 9$, those not exceeding 4 appeared $h = 5089$ times.

1.15. Simulate a sample of size $n = 1000$ from Bernoulli's distribution $Bi(1, 3/5)$ and check, as in Problem 1.13, whether the experimental data correspond to the theoretical prediction.

| *Hint.* Use Problem 1.1.

1.16. Suppose that an experiment consists of tossing 12 dice. The observable random variable ξ is equal to the number of dice with a 4, a 5, or a 6. Let h_i be the number of trials in which the values $\xi = i$, $i = 0, 1, \ldots, 12$, were observed. The data for $n = 4096$ trials are given [8] in the following table:

i	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
h_i	0	7	60	198	430	731	948	847	536	257	71	11	0	$n = 4096$

(a) Construct the frequency graph h_i/n and compare it with the graph of the function $ce^{-x^2/2}$.

(b) Compute the sample mean and variance, the skewness coefficient, and kurtosis.

(c) Assume that $\mathcal{L}(\xi) = Bi(12, 1/2)$ and find δ from the condition $\mathbf{P}(|X - \alpha_1| \leqslant \delta) = 0.998$. Compare δ with the deviation of the sample mean from the theoretical α_1 as calculated from the given data.

|*Hint.* When estimating the probability in (c), use the theorem on the asymptotic normality of a sample mean.

1.17. (Continued from Problem 1.16.) Let the random variable ξ of the previous experiment be equal to the number of dice with a 6. The observed data are tabulated [8] as

i	0	1	2	3	4	5	6	$\geqslant 7$	Total
h_i	447	1145	1181	796	380	115	24	8	$n = 4096$

Answer the questions of Problem 1.16 if $\mathcal{L}(\xi) = Bi(12, 1/6)$.

1.18. Simulate a sample of size $n = 1000$ from the distribution $\mathcal{L}(\xi) = Bi(4, 1/3)$ and analyze the obtained data as in Problem 1.16.

|*Hint.* Use Problem 1.12.

1.19. Suppose that we observe 500 randomly chosen watches in shop windows. Let i be the number of the interval between the ith and $(i + 1)$th hours, $i = 0, 1, \ldots, 11$, and let h_i be the number of watches indicating the ith interval. The observation results are grouped [3] in the following table:

i	0	1	2	3	4	5	6	7	8	9	10	11	Total
h_i	41	34	54	39	49	45	41	33	37	41	47	39	$n = 500$

(a) Construct a frequency polygon and compare it with the plot of the function $f(x) = c$, $0 \leqslant x \leqslant 12$.

(b) Assume that these data are independent observations on a discrete random variable ξ whose values coincide with the middle points of the respective intervals (i.e., 0.5, 1.5, ..., 11.5) and calculate the sample mean and variance.

(c) Assuming that the random variable ξ from (b) has a uniform distribution, find δ from the condition $\mathbf{P}(|\bar{X} - \alpha_1| \leqslant \delta) = 0.98$ and compare it with the observed deviation $|\bar{X} - \alpha_1|$.

1.20. Simulate a sample from the polynomial distribution $M(500; 1/5, 1/5, 1/5, 1/5, 1/5)$ and, assuming that these data are observations on a random variable ξ having the values $-2, -1, 0, 1, 2$, analyze the respective data as it was done in Problem 1.19.

| *Hint.* Use Problem 1.3.

1.21. Suppose that we observed a non-negative continuous random variable ξ. Its values (rounded to 0.01 and placed in the order of magnitude) for $n = 50$ trials were 0.01, 0.01, 0.04, 0.17, 0.18, 0.22, 0.22, 0.25, 0.25, 0.29, 0.42, 0.46, 0.47, 0.47, 0.56, 0.59, 0.67, 0.68, 0.70, 0.72, 0.76, 0.78, 0.83, 0.85, 0.87, 0.93, 1.00, 1.01, 1.01, 1.02, 1.03, 1.05, 1.32, 1.34, 1.37, 1.47, 1.50, 1.52, 1.54, 1.59, 1.71, 1.90, 2.10, 2.35, 2.46, 2.46, 2.50, 3.73, 4.07, 6.03. Construct an empirical distribution function and a histogram. Compare the histogram with the graph of the function $ce^{-x/a}$, $x > 0$. Compute the sample mean and variance.

1.22. Suppose that a sample of size $n = 100$ was 0.144, 0.937, 1.787, −1.052, −0.192, 0.169, 2.623, 2.135, 1.759, 0.811, 0.724, −0.110, 1.752, −0.378, 0.417, 1.360, 1.365, 2.587, 1.621, 2.344, 1.379, 0.560, 1.858, 2.453, −0.356, 1.503, −0.134, 2.950, −0.816, 0.717, 2.468, 1.131, 1.047, 1.355, 1.162, −0.491, 0.261, −0.183, 0.467, 0.502, −0.805, 0.228, 2.286, 0.364, −0.312, −0.045, 2.559, 0.129, 0.898, 0.877, 3.285, 1.554, 1.418, 0.423, −0.489, −0.255, 1.092, 0.402, −0.051, 0.020, 0.398, 1.399, 2.121, −0.026, 1.087, 2.018, −0.437, 1.661, 1.091, 0.363, 1.229, 0.416, 1.705, 1.124, 1.341, 2.320, 0.176, −0.541, 0.837, 3.329, 2.382, −0.454, 2.537, −0.299, 1.363, 0.644, 0.975, 1.294, 3.194, 0.605, 1.978, 1.109, 2.434, −0.094, 0.735, 0.143, −0.421, −0.773, 1.570, 0.947. Construct an empirical distribution function and a histogram. Calculate the sample mean and variance, the skewness coefficient, and kurtosis.

1.23. Let α-particles be radiated by a radioactive substance during 7.5 seconds. Suppose that the following data were obtained in $n = 2608$ experiments (h_i is the number of trials for which the number of particles $\xi = i$, $i = 0, 1, \ldots$), i.e.,

i	0	1	2	3	4	5	6	7	8	9	10	11	$\geqslant 12$	Total
h_i	57	203	383	525	532	408	273	139	45	27	10	4	2	$n = 2608$

Construct the frequency graph h_i/n and calculate the sample mean and variance [5].

In the problems below $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample from a distribution $\mathscr{L}(\xi)$, and $F(x)$ and $F_n(x)$ are the theoretical and empirical distribution functions, respectively (see (1.1)).

1.24. Given a point x_0, such that $0 < F(x_0) < 1$, and a number t, estimate the probability of the event

$$|F_n(x) - F(x_0)| \leqslant t/\sqrt{n}$$

for large n.

[*Hint.* Use the De Moivre-Laplace theorem.

1.25. Let $x_1 < x_2$ be two given points on a number line, such that $0 < F(x_1) \leqslant F(x_2) < 1$. Prove that

$$\operatorname{cov}(F_n(x_1), \quad F_n(x_2)) = \frac{1}{n} F(x_1)(1 - F(x_2)).$$

Hint. Represent the random variables $\mu_n(x_1)$ and $\Delta_n(x_1, x_2) = \mu_n(x_2) - \mu_n(x_1)$ as the sums of independent indicators

$$\mu_n(x_1) = \eta_1 + \ldots + \eta_n,$$

where $\eta_i = I(X_i \leqslant x_1)$, $i = 1, \ldots, n$, and

$$\Delta_n(x_1, x_2) = \zeta_1 + \ldots + \zeta_n,$$

where $\zeta_j = I(x_1 < X_j \leqslant x_2)$, $j = 1, \ldots, n$.

1.26. Let $x_1 < x_2 < \ldots < x_{N-1}$ be given points on a number line, such that $0 < F(x_1) < F(x_2) < \ldots < F(x_{N-1}) < 1$. Examine the random variables $\nu_i = \mu_n(x_i) - \mu_n(x_{i-1})$, $i = 1, \ldots, N$ (here $\mu_n(x_0) = 0$, $\mu_n(x_N) = n$), and make sure that the random vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ has a polynomial distribution $M(n; p_1, \ldots, p_N)$, where $p_i = F(x_i) - F(x_{i-1})$, $i = 1, \ldots, N$, $F(x_0) = 0$, $F(x_N) = 1$. Derive the result obtained in Problem 1.25.

1.27. Derive the formulas

$$\mathbf{E}A_{nk} = \mathbf{E}\xi^k = \alpha_k, \quad \operatorname{cov}(A_{nk}, A_{ns}) = \frac{\alpha_{k+s} - \alpha_k\alpha_s}{n},$$

$$\mathbf{E}S^2 = \frac{n-1}{n}\mu_2, \quad \mathbf{D}S^2 = \frac{(n-1)^2}{n^3}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right),$$

$$\mu_k = \mathbf{E}(\xi - \alpha_1)^k, \quad \operatorname{cov}(\overline{X}, S^2) = \frac{n-1}{n^2}\mu_3$$

for the moments of sampling moments. Calculate the moments for $\mathscr{L}(\xi) = \mathscr{N}(\mu, \sigma^2)$.

1.28. Prove that for any fixed $r \geqslant 2$, $1 \leqslant k_1 < \ldots < k_r$, the joint distribution of the sampling moments $A_{nk_1}, \ldots, A_{nk_r}$ as $n \to \infty$ is asymptotically normal as $\mathcal{N}\left(\alpha = (\alpha_{k_1}, \ldots, \alpha_{k_r}), \frac{1}{n}\Sigma\right)$, where $\Sigma = \|\sigma_{ij} = \alpha_{k_i+k_j} - \alpha_{k_i}\alpha_{k_j}\|_1^r$, i.e., $\mathcal{L}(\sqrt{n}(A_{nk_i} - \alpha_{k_i}), i = 1, \ldots, r) \to \mathcal{N}(\mathbf{0}, \Sigma)$ (it is assumed that all the theoretical moments exist). Besides, if $\varphi(\mathbf{x})$, $\mathbf{x} = (x_1, \ldots, x_r)$, is any differentiable function, then $\mathcal{L}(\sqrt{n}(\varphi(A_{nk_1}, \ldots, A_{nk_r}) - \varphi(\alpha))) \to \mathcal{N}(0, v^2)$ under the condition that $v \neq 0$, where

$$v^2 = \mathbf{b}'\Sigma\mathbf{b}, \quad \mathbf{b} = \left(\frac{\partial\varphi(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial\varphi(\mathbf{x})}{\partial x_r}\right)\bigg|_{\mathbf{x}=\alpha}.$$

Hint. Apply the Central Limit Theorem for vector random variables and the assertion 4° from Sec. 1.4.

1.29*. Prove that for $n \to \infty$ the sample variance S_n^2 is asymptotically normal as $\mathcal{N}(\mu_2, (\mu_4 - \mu_2^2)/n)$ and then $\mathbf{E}S_n^2 \sim \mu_2$, $\mathbf{D}S_n^2 \sim (\mu_4 - \mu_2^2)/n$ (we assume that $\mu_4 < \infty$).

Hint. Use the assertion 2° (a) from Sec. 1.4.

1.30. Prove that the joint distribution function of two order statistics $X_{(r)}$ and $X_{(s)}$, $1 \leqslant r < s \leqslant n$, has the form

$$F_{r,s}(x_1, x_2) = \sum_{m=r}^{n} \sum_{j=\max(0,\, s-m)}^{n-m} \frac{n!}{m!j!(n-m-j)!} F^m(x_1) \times (F(x_2) - F(x_1))^j(1 - F(x_2))^{n-m-j}$$

if $x_1 < x_2$, and the form

$$F_{r,s}(x_1, x_2) = \mathbf{P}(X_{(s)} \leqslant x_2) = F_s(x_2) = \sum_{j=s}^{n} C_n^j F^j(x_2)(1 - F(x_2))^{n-j}$$

if $x_1 \geqslant x_2$. Using this, derive the formula for $F_r(x) = \mathbf{P}(X_{(r)} \leqslant x)$.

1.31. Let the distribution $\mathcal{L}(\xi)$ be absolutely continuous and its density be $F'(x) = f(x)$. Derive the formula

$$g_{k_1 \ldots k_r}(x_1, \ldots, x_r) = \frac{n!}{(k_1-1)!(k_2-k_1-1)!\ldots(k_r-k_{r-1}-1)!(n-k_r)!} \times F^{k_1-1}(x_1)(F(x_2) - F(x_1))^{k_2-k_1-1} \ldots \times (F(x_r) - F(x_{r-1}))^{k_r-k_{r-1}-1}(1 - F(x_r))^{n-k_r} f(x_1)\ldots f(x_r),$$

$$x_1 < x_2 < \ldots < x_r$$

for the density of the joint distribution of the order statistics $X_{(k_1)}, \ldots, X_{(k_r)}$, $1 \leqslant k_1 < \ldots < k_r \leqslant n$. Specifically, the joint density of all the n order statistics $X_{(1)}, \ldots, X_{(n)}$ is

$$g_{1\ldots n}(x_1, \ldots, x_n) = n! f(x_1) \ldots f(x_n), \quad x_1 < x_2 < \ldots < x_n.$$

1.32*. Prove that if in some neighbourhoods of the quantiles ζ_{p_1} and ζ_{p_2}, $0 < p_1 < p_2 < 1$, the density $f(x)$ is continuous together with its derivative and $f(\zeta_{p_i}) > 0$, $i = 1, 2$, then as $n \to \infty$ the sample quantiles $Z_{n, p_i} = X_{([np_i] + 1)}$, $i = 1, 2$, are asymptotically normal as $\mathcal{N}\left((\zeta_{p_1}, \zeta_{p_2}), \frac{1}{n} \|\sigma_{ij}\|_1^2\right)$, where $\sigma_{ij} = \dfrac{p_i(1 - p_j)}{f(\zeta_{p_i}) f(\zeta_{p_j})}$, $i \leqslant j$. Generalize the proof to the case of r-quantiles.

1.33*. Prove that for a sample from an absolutely continuous distribution the extreme order statistics $X_{(r)}$ and $X_{(n-s+1)}$ are asymptotically independent as $n \to \infty$ for fixed r, $s \geqslant 1$.

Hint. Go over to the random variables $\varkappa_n = nF(X_{(r)})$ and $\eta_n = n[1 - F(X_{(n-s+1)})]$ and use the result obtained in Problem 1.31.

1.34. Let $\mathcal{L}(\xi) = \Gamma(1, 1)$. Prove that the random variables $Y_r = (n - r + 1)(X_{(r)} - X_{(r-1)})$, $r = 1, \ldots, n$, $X_{(0)} = 0$, are independent and similarly distributed with the density $f(x) = e^{-x}$, $x > 0$. Calculate $\mathbf{E}X_{(k)}$, $\mathbf{D}X_{(k)}$ and investigate the asymptotic behaviour of $\mathbf{E}X_{(n)}$ and $\mathbf{D}X_{(n)}$ as $n \to \infty$.

Hint. Use the formula $\sum\limits_{r=1}^{n} x_r = \sum\limits_{r=1}^{n} (n - r + 1)(x_r - x_{r-1})$, $x_0 = 0$, and the result of Problem 1.31.

1.35. Make sure that in the case of $\mathcal{L}(\xi) = R(0, 1)$ the distributions of the order statistics have the form

$$\mathcal{L}(X_{(k)}) = B(k, n - k + 1),$$
$$\mathcal{L}(X_{(l)} - X_{(k)}) = B(l - k, n - l + k + 1),$$

$1 \leqslant k < l \leqslant n$. Calculate the means and variances of these distributions and $\operatorname{cov}(X_{(k)}, X_{(l)})$.

1.36. Let $\mathcal{L}(\xi) = R(a, b)$. Prove that the density of the joint distribution of the extreme values $X_{(1)}$ and $X_{(n)}$ of a sample has the form $\dfrac{n(n - 1)}{(b - a)^n} (x_2 - x_1)^{n-2}$, $a \leqslant x_1 \leqslant x_2 \leqslant b$. Derive the formulas

$$\mathbf{E}X_{(1)} = \frac{na + b}{n + 1}, \quad \mathbf{E}X_{(n)} = \frac{a + nb}{n + 1},$$

$$\mathbf{D}X_{(1)} = \mathbf{D}X_{(n)} = \frac{n(b-a)^2}{(n+1)^2(n+2)},$$

$$\operatorname{cov}(X_{(1)}, X_{(n)}) = \frac{(b-a)^2}{(n+1)^2(n+2)}.$$

1.37. Let $\mathscr{L}(\xi)$ be Weibull's distribution $W(a, \alpha, b)$. Find the distribution of the minimum value of the sample $X_{(1)}$ and compute $\mathbf{E}X_{(1)}$ and $\mathbf{D}X_{(1)}$.

1.38. Let $\mathbf{X}_i = (X_{i1}, X_{i2})$, $i = 1, \ldots, n$, be independent observations on a two-dimensional random variable $\boldsymbol{\xi} = (\xi_1, \xi_2)$ with the distribution function $F(x_1, x_2)$. The empirical distribution function is

$$F_n(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n} I(X_{i1} \leqslant x_1) I(X_{i2} \leqslant x_2)$$

(compare it with (1.1)). Calculate $\mathbf{E}F_n(x_1, x_2)$ and $\mathbf{D}F_n(x_1, x_2)$ and show that $F_n(x_1, x_2) \to F(x_1, x_2)$ as $n \to \infty$. Construct the sample correlation coefficient ϱ_n and show that $\varrho_n \xrightarrow{\mathbf{P}} \varrho = \operatorname{corr}(\xi_1, \xi_2)$ if $\mathbf{E}(\xi_1^2\xi_2^2) < \infty$ and $\mathbf{D}\xi_j > 0$, $j = 1, 2$.

1.39. A distribution $\mathscr{L}_a$ which depends on a parameter a is said to be *reproducible* in this parameter if the independent random variables ξ_1 and ξ_2 distributed as $\mathscr{L}_{a_1}$ and $\mathscr{L}_{a_2}$, respectively, satisfy the condition $\mathscr{L}(\xi_1 + \xi_2) = \mathscr{L}_{a_1+a_2}$ (this is sometimes written as $\mathscr{L}_{a_1} * \mathscr{L}_{a_2} = \mathscr{L}_{a_1+a_2}$, where $*$ stands for convolution).

Make sure that the following assertions are true:

(1) $\mathscr{N}(\mu_1, \sigma_1^2) * \mathscr{N}(\mu_2, \sigma_2^2) = \mathscr{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$;

(2) $\Gamma(a, \lambda_1) * \Gamma(a, \lambda_2) = \Gamma(a, \lambda_1 + \lambda_2)$;

(3) $M(n_1; p_1, \ldots, p_N) * M(n_2; p_1, \ldots, p_N) = M(n_1 + n_2; p_1, \ldots, p_N)$; specifically, $Bi(n_1, p) * Bi(n_2, p) = Bi(n_1 + n_2, p)$;

(4) $\Pi(\lambda_1) * \Pi(\lambda_2) = \Pi(\lambda_1 + \lambda_2)$;

(5) $\overline{Bi}(r_1, p) * \overline{Bi}(r_2, p) = \overline{Bi}(r_1 + r_2, p)$.

Hint. Use the fact that the characteristic function of a sum of independent random variables is equal to the product of the characteristic functions of the terms. If the random variables are discrete, it is more convenient to use the generating functions $\mathbf{E}x^{\xi}$ instead of the characteristic functions $\mathbf{E}e^{it\xi}$.

1.40*. Suppose that a random vector $\mathbf{Y}$ has a non-degenerate normal distribution $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $Q = (\mathbf{Y} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})$, where the matrix $\mathbf{A}$ satisfies the condition $\mathbf{A} = \mathbf{A\Sigma A}$. Prove that $\mathscr{L}(Q) = \chi^2(m)$, where $m = \operatorname{tr}(\mathbf{A\Sigma})$.

In the case of $\mathbf{A} = \mathbf{\Sigma}^{-1}$ the number of the degrees of freedom m coincides with the dimensionality of the vector $\mathbf{Y}$.

1.41. Let the joint distribution of two random variables X and Y be such that the conditional distribution of X when $Y = y$ is $\mathcal{N}(y, \sigma_1^2)$-normal, while $\mathcal{L}(Y) = \mathcal{N}(\mu, \sigma_2^2)$. Prove that $\mathcal{L}(X) = \mathcal{N}(\mu, \sigma_1^2 + \sigma_2^2)$.

Hint. Compute the distribution density of X by the formula

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) f_Y(y)\, dy,$$

where $f_{X,Y}(x, y)$ is the density of the conditional distribution $\mathcal{L}(X \mid Y = y)$.

1.42. Suppose that the random variables X_1 and X_2 are independent and

$$\mathcal{L}(X_1) = \Gamma(a, \lambda), \quad \mathcal{L}(X_1 + X_2) = \Gamma(a, \lambda + \mu), \quad \mu > 0.$$

How is the random variable X_2 distributed?

Hint. Calculate the characteristic function for X_2 (see the solution to Problem 1.39 (2)).

1.43. Let ξ_1 and ξ_2 be independent random variables uniformly distributed on the segment $[0, 1]$. Show that the quantities $\eta_1 = \sqrt{-2 \ln \xi_2} \cos (2\pi\xi_1)$ and $\eta_2 = \sqrt{-2 \ln \xi_2} \sin (2\pi\xi_1)$ are independent and normally distributed with the parameters $(0, 1)$.

Hint. Use formula (1.2).

1.44. Let the random variables X_1 and X_2 be independent, and let $\mathcal{L}(X_i) = \Gamma(a, \lambda_i)$, $i = 1, 2$. Prove that the random variables $Y_1 = X_1 + X_2$ and $Y_2 = X_1/(X_1 + X_2)$ are independent, and $\mathcal{L}(Y_1) = \Gamma(a, \lambda_1 + \lambda_2)$ is the reproducibility in λ (see Problem 1.39 (2)), $\mathcal{L}(Y_2) = B(\lambda_1, \lambda_2)$.

1.45. Prove that $\mathcal{L}\left(\dfrac{\chi_n^2 - n}{\sqrt{2n}}\right) \to \mathcal{N}(0, 1)$ as $n \to \infty$ and $\mathbf{E}(\chi_n^2)^k = n(n + 2)\ldots(n + 2(k - 1))$, $k = 1, 2, \ldots$.

Hint. Use the reproducibility of the distribution $\Gamma(a, \lambda)$ with respect to λ and apply the Central Limit Theorem.

1.46. Show that $\mathcal{L}\left(a + \dfrac{\xi}{\eta}\right) = C(a)$, where the random variables ξ and η are independent and $\mathcal{N}(0, \sigma^2)$-normal, and

$$\mathcal{L}(a + \tan \xi) = C(a),$$

where $\mathcal{L}(\xi) = R\left(-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right)$.

1.47. Show that if $\mathscr{L}(t_n) = S(n)$, then the moments $\mathbf{E}t_n^k$ exist if and only if $k < n$ and are of the form

$$\mathbf{E}t_n^{2r} = \frac{1 \times 3 \times \ldots \times (2r-1)n^r}{(n-2)(n-4)\ldots(n-2r)}, \quad 2r < n,$$

$$\mathbf{E}t_n^{2r+1} = 0, \quad 2r + 1 < n.$$

Prove that $S(n) \to \mathscr{N}(0, 1)$ as $n \to \infty$, and, moreover, the density $s_n(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Show that $\mathscr{L}\left(\frac{1}{1 + t_n^2/n}\right) = B\left(\frac{n}{2}, \frac{1}{2}\right)$.

Hint. Use Stirling's formula for the gamma function $\Gamma(z) \sim \sqrt{2\pi z}\, z^{z-1} e^{-z}$, $z \to \infty$, and apply the law of large numbers to the random variable χ_n^2/n (see the solution to Problem 1.45). When calculating the moments, take into account that $t_n = \eta \sqrt{\frac{n}{\chi_n^2}}$ and the terms are independent. Use Problem 1.44.

1.48*. Let $F(x; n_1, n_2)$ be the distribution function of the Snedecor law $S(n_1, n_2)$, and let $\mathrm{B}(x; a, b)$ be the distribution function for $B(a, b)$. Show that

$$F(x; n_1, n_2) = \mathrm{B}\left(\frac{n_1 x}{n_2 + n_1 x}; \frac{n_1}{2}, \frac{n_2}{2}\right), \quad x > 0.$$

Derive the expression for the density $f_{n_1, n_2}(x)$ of the distribution $S(n_1, n_2)$. Find the distribution moments.

1.49*. (Continued from Problem 1.48.) Prove that for any fixed $x \in (0, 1)$ and $a > 0$

$$\lim_{b \to \infty} \frac{1}{b} \ln [1 - \mathrm{B}(x; a, b)] = \ln (1 - x).$$

Show that for any fixed $t > 0$ and $m \geqslant 1$

$$\lim_{n \to \infty} \frac{1}{n} \ln [1 - F(tn; m, n)] = -\frac{1}{2} \ln (1 + mt).$$

Hint. Use Stirling's formula (see the hint to Problem 1.47) and the theorem

$$\int_x^1 u^{a-1}(1-u)^{b-1}\, du = c^{a-1} \int_x^1 (1-u)^{b-1}\, du = \frac{c^{a-1}}{b}(1-x)^b, \quad c \in [x, 1],$$

about the mean.

1.50*. Show that the density $s_n(x)$ of Student's distribution $S(n)$ can be expressed through the density $f_{1,n}(x)$ of Snedecor's distribution $S(1, n)$ as

$$s_n(x) = |x| f_{1,n}(x^2).$$

Prove the relation

$$\lim_{n \to \infty} \frac{1}{n} \ln \mathbf{P}(t_n > d\sqrt{n}) = -\frac{1}{2} \ln (1 + d^2) \quad \forall d > 0.$$

| *Hint.* Use the fact that $\mathscr{L}(t_n^2) = S(1, n)$.

1.51. Let $\mathbf{X} = (X_1, \ldots, X_l, X_{l+1}, \ldots, X_{l+m})$ be a sample from the exponential distribution $\mathscr{L}(\xi) = \Gamma(a, 1)$. Investigate the random variable $Y = \dfrac{m}{l} \dfrac{X_1 + \ldots + X_l}{X_{l+1} + \ldots + X_{l+m}}$ and prove that $\mathscr{L}(Y) = S(2l, 2m)$.

| *Hint.* Use the fact that $\mathscr{L}(t_n^2) = S(1, n)$.

1.52. Let an integer-valued random vector $\nu = (\nu_1, \ldots, \nu_N)$ have the polynomial distribution $M(n; p_1, \ldots, p_N)$.

(a) Show that the generating function for $(\nu_1, \ldots, \nu_k)$, $k \leqslant N$, has the form

$$\mathbf{E}(x_1^{\nu_1} \ldots x_k^{\nu_k}) = \left[1 + \sum_{i=1}^{k} p_i(x_i - 1)\right]^n,$$

and, specifically, $\mathscr{L}(\nu_1) = Bi(n, p_1)$.

(b) Derive the general formula

$$\mathbf{E}(\nu_1)_{k_1} \ldots (\nu_N)_{k_N} = (n)_{k_1 + \ldots + k_N} p_1^{k_1} \ldots p_N^{k_N}$$

for mixed factorial moments.

(c) Let $\eta^j = \sum_{i=1}^{N} c_i^j \nu_i$, $\bar{c}^j = \sum_{i=1}^{N} c_i^j p_i$, $j = 1, 2$, $\overline{c^1 c^2} = \sum_{i=1}^{N} c_i^1 c_i^2 p_i$.

Show that

$$\mathbf{E}\eta^j = n\bar{c}^j, \ \operatorname{cov}(\eta^1, \eta^2) = n(\overline{c^1 c^2} - \bar{c}^1 \bar{c}^2).$$

| *Hint.* Use the result obtained in Problem 1.39 (3).

1.53*. (Continued from Problem 1.52.) Prove that as $n \to \infty$ for any $k < N$ and fixed $p_i \in (0, 1)$, $i = 1, \ldots, N$,

$$\mathscr{L}((\nu_j - np_j)/\sqrt{n}, \ j = 1, \ldots, k) \to \mathscr{N}(\mathbf{0}, \Sigma_k = \|\sigma_{ij}\|_1^k),$$

where

$$\sigma_{ij} = \begin{cases} p_i(1 - p_i) & \text{for} \quad i = j \\ -p_i p_j & \text{for} \quad i \neq j \end{cases} \quad \text{and} \quad |\Sigma_k| \neq 0.$$

Hint. Use the theorem on the continuity of characteristic functions.

1.54. Prove that if the random variables $\xi_1, \ldots, \xi_N$ are independent and $\mathscr{L}(\xi_j) = \Pi(\lambda_j)$, $j = 1, \ldots, N$, then the conditional distribution is

$$\mathscr{L}(\xi_1, \ldots, \xi_N | \xi_1 + \ldots + \xi_N = n) = M(n; p_1, \ldots, p_N),$$

where $p_j = \dfrac{\lambda_j}{\lambda_1 + \ldots + \lambda_N}$, $j = 1, \ldots, N$. It follows from here that $\mathscr{L}(\xi_1 | \xi_1 + \ldots + \xi_N = n) = Bi(n, p_1)$.

1.55. Suppose that we have two random variables ξ and Λ and $\mathscr{L}(\Lambda) = \Gamma(a, r)$ for some $a > 0$ and integer $r \geqslant 1$, and we also have a conditional distribution $\mathscr{L}(\xi | \Lambda = \lambda) = \Pi(\lambda)$. Show that the unconditional distribution $\mathscr{L}(\xi) = \overline{Bi}(r, p)$ for $p = a/(a + 1)$.

1.56. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from $\mathscr{N}(\mu, \sigma^2)$. Prove that $\overline{X}$ and $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$ are independent. Prove that the sample mean $\overline{X}$ and variance S^2 are independent.

Hint. Use the fact that the uncorrelatedness of normal random variables implies their independence.

1.57*. Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample from $\mathscr{N}(0, 1)$, and the quadratic form $Q = \mathbf{X}'\mathbf{X}$ is represented as the sum $Q = Q_1 + Q_2$ of two quadratic forms, where $Q_i = \mathbf{X}'\mathbf{A}_i\mathbf{X}$ and rank $\mathbf{A}_i = n_i$, $i = 1, 2$. Prove that if $n_1 + n_2 = n$, then Q_1 and Q_2 are independent and $\mathscr{L}(Q_i) = \chi^2(n_i)$, $i = 1, 2$.

Hint. Check whether the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are idempotent and $\mathbf{A}_1\mathbf{A}_2 = \mathbf{0}$. Then use the assertions 1° and 2° from Sec. 1.6 (1).

Remark. A stronger assertion is true, i.e., if $Q = Q_1 + \ldots + Q_k$, where $Q_i = \mathbf{X}'\mathbf{A}_i\mathbf{X}$, rank $\mathbf{A}_i = n_i$, $i = 1, \ldots, k$, then $n = n_1 + \ldots + n_k \Leftrightarrow Q_1, \ldots, Q_k$ are independent and $\mathscr{L}(Q_i) = \chi^2(n_i)$, $i = 1, \ldots, k$.

1.58*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the normal distribution $\mathscr{N}(\mu, \sigma^2)$. Find the distribution of the random variable $\eta = \dfrac{X_1 - \overline{X}}{\sqrt{n-1}S}$.

1.59*. Use the notations of Problem 1.38 and assume that

$$\mathscr{L}(\xi) = \mathscr{N}\left((\mu_1, \mu_2), \quad \Sigma = \begin{Vmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho \\ \sigma_1\sigma_2\varrho & \sigma_2^2 \end{Vmatrix}\right), \quad -1 < \varrho < 1.$$

(a) Prove that $(\overline{X_1}, \overline{X_2})$ and (S_1^2, S_{12}, S_2^2) are independent.

(b) Assume that $Q = n(\overline{X_1} - \mu_1, \overline{X_2} - \mu_2)'\Sigma^{-1}(\overline{X_1} - \mu_1, \overline{X_2} - \mu_2)$ and prove that $\mathscr{L}(Q) = \chi^2(2)$.

(c) Let $n > 2$ and $T = \sqrt{n-2}\,\varrho_n/\sqrt{1-\varrho_n^2}$, where $\varrho_n = S_{12}/S_1S_2$ is the sample correlation coefficient. Prove that for $\varrho = 0$ we have

$$\mathscr{L}(T) = S(n-2)$$

and find the distribution for ϱ_n.

Hints. (a) See Problem 1.56 and its solution.

(b) Use Problem 1.40.

(c) Use the fact [3] that the density of the joint distribution of the random variables (S_1^2, S_{12}, S_2^2) has the form (for $\varrho = 0$)

$$f(x_1, x_{12}, x_2) = \frac{n^{n-1}(x_1x_2 - x_{12}^2)^{(n-4)/2}}{4\pi\Gamma(n-2)(\sigma_1\sigma_2)^{n-1}}\, e^{-\frac{n}{2}\left(\frac{x_1}{\sigma_1^2}+\frac{x_2}{\sigma_2^2}\right)},$$

$$x_1, x_2 > 0; \quad x_{12}^2 < x_1x_2.$$

Then consider new random variables

$$Y_1 = \frac{\sqrt{n}}{\sigma_1}\frac{S_{12}}{S_2}, \quad Y_2 = \frac{n}{\sigma_1^2}\left(S_1^2 - \frac{S_{12}^2}{S_2^2}\right), \quad Y_3 = \frac{n}{\sigma_2^2}S_2^2$$

and take into account that $T = Y_1/\sqrt{Y_2/(n-2)}$.

1.60*. Let $\overline{X}$ and S^2 be the sample mean and variance for a sample of size n from the distribution $\Pi(\lambda)$. Prove that

$$\mathscr{L}\left(T_n \equiv \sqrt{\frac{n-1}{2}}\left(\overline{X} - \frac{n}{n-1}S^2\right)\Big/\overline{X}\right) \to \mathscr{N}(0, 1)$$

as $n \to \infty$ and for any $\lambda > 0$.

Hint. Use Problems 1.27-29. First show the asymptotic normality $\mathscr{N}(0, 1)$ of the random variable $\zeta_n = \sqrt{\frac{n-1}{2}} \times \left(\overline{X} - \frac{n}{n-1}S^2\right)\Big/\lambda$ and then use the fact that $\overline{X}/\lambda \xrightarrow{P} 1$ as $n \to \infty$ and the assertion 2° from Sec. 1.4. When computing the moments, use the formulas

$$\mu_2 = \mu_3 = \lambda, \quad \mu_4 = \lambda + 3\lambda^2$$

for central moments of Poisson's distribution $\Pi(\lambda)$.

1.61. Let $X_1, \ldots, X_n$ be independent observations on a random variable ξ with $\mathbf{E}\xi = \mu$, $\mathbf{D}\xi = \sigma^2 > 0$, $\mathbf{E}\xi^4 < \infty$, and let $\overline{X}$ and S^2 be the respective sample mean and variance. Prove that for $n \to \infty$ we have

$$\mathscr{L}(T_n \equiv \sqrt{n}(\overline{X} - \mu)/S) \to \mathscr{N}(0, 1).$$

Hint. Use the Central Limit Theorem, the convergence $S/\sigma \xrightarrow{P} 1$ as $n \to \infty$, and the assertion 2° (c) from Sec. 1.4.

1.62. Suppose that $\mathscr{L}(\xi_1, \xi_2) = \mathscr{N}\left((\mu_1, \mu_2), \begin{Vmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho \\ \sigma_1\sigma_2\varrho & \sigma_2^2 \end{Vmatrix}\right)$, where $-1 < \varrho = \operatorname{corr}(\xi_1, \xi_2) < 1$. Prove that the conditional distribution is $\mathscr{L}(\xi_2 | \xi_1 = x) = \mathscr{N}(m(x), \sigma^2)$, where the conditional mean

$$m(x) = \mathbf{E}(\xi_2 | \xi_1 = x) = \mu_2 + \frac{\sigma_2}{\sigma_1}\varrho(x - \mu_1)$$

is a regression function of ξ_2 on ξ_1, which in our case is a linear function in x, and the conditional variance $\sigma^2 = \mathbf{D}(\xi_2 | \xi_1 = x) = \sigma_2^2(1 - \varrho^2)$ does not depend on x. Check whether there are other distributions of the form $\mathscr{L}(\xi_1, \xi_2)$ which are not normal but have the same properties as the conditional distribution $\mathscr{L}(\xi_2 | \xi_1 = x)$.

Hints. (1) Calculate the conditional density ξ_2 under the condition $\xi_1 = x$ using the formula $f_{\xi_2|\xi_1}(y | x) = f_{\xi_1\xi_2}(x, y)/f_{\xi_1}(x)$ and make sure that it is equal to $\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y - m(x))^2}{2\sigma^2}\right\}$ with $m(x)$ and σ^2 as given above.
(2) Consider the joint distribution density $f_{\xi_1\xi_2}(x, y) = f_{\xi_1}(x) f_{\xi_2|\xi_1}(y|x)$, where $f_{\xi_2|\xi_1}(y|x)$ is the conditional and $f_{\xi_1}(x)$ the unconditional distribution density.

1.63*. (A generalized variant of Problem 1.62.) Let $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ be random vectors of arbitrary dimensionality, $\mathbf{E}(\mathbf{X}^{(i)}) = \mu^{(i)}$, $\mathbf{D}(\mathbf{X}^{(i)}) = \Sigma_{ii}$, $i = 1, 2$, $\operatorname{cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \Sigma_{12}$, $\operatorname{cov}(\mathbf{X}^{(2)}, \mathbf{X}^{(1)}) = \Sigma_{21}$ $(= \Sigma_{12}')$, $\mathscr{L}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \mathscr{N}\left((\mu^{(1)}, \mu^{(2)}), \Sigma = \begin{Vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{Vmatrix}\right)$, $|\Sigma| \neq 0$. Prove that the conditional distribution $\mathscr{L}(\mathbf{X}^{(2)} | \mathbf{X}^{(1)} = \mathbf{x}^{(1)}) = \mathscr{N}(M(\mathbf{x}^{(1)}), \mathbf{B})$, where

$$M(\mathbf{x}^{(1)}) = \mu^{(2)} + \mathbf{A}(\mathbf{x}^{(1)} - \mu^{(1)}), \quad \mathbf{A} = \Sigma_{21}\Sigma_{11}^{-1},$$
$$\mathbf{B} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Show that when $\dim \mathbf{X}^{(2)} = 1$, $\dim \mathbf{X}^{(1)} = p - 1$, we have

$$\mathbf{A} = -\frac{1}{\sigma^{pp}}(\sigma^{1p}, \sigma^{2p}, \ldots, \sigma^{p-1,p}), \quad \mathbf{B} = \frac{1}{\sigma^{pp}},$$

where $\Sigma^{-1} = |\sigma^{ij}|_1^p$.

Hints. (1) Consider a linear transformation $\mathbf{Y}^{(1)} = \mathbf{X}^{(1)}$, $\mathbf{Y}^{(2)} = \mathbf{X}^{(2)} - \mathbf{A}\mathbf{X}^{(1)}$ and make sure that $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are in-

dependent. Then

$$\mathscr{L}(\mathbf{X}^{(2)}|\mathbf{X}^{(1)} = \mathbf{x}^{(1)}) = \mathscr{L}(\mathbf{Y}^{(2)} + \mathbf{A}\mathbf{Y}^{(1)}|\mathbf{Y}^{(1)} = \mathbf{x}^{(1)})$$
$$= \mathscr{L}(\mathbf{Y}^{(2)} + \mathbf{A}\mathbf{x}^{(1)}).$$

(2) Write the joint density

$$f_{\mathbf{X}^{(1)}\mathbf{X}^{(2)}}(x_1, \ldots, x_{p-1}, x_p)$$
$$= C \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{p}(x_i - \mu_i)(x_j - \mu_j)\sigma^{ij}\right\}$$
$$= C \exp\left\{-\frac{1}{2}\left[(x_p - \mu_p)^2\sigma^{pp} + 2(x_p - \mu_p)\sum_{i=1}^{p-1}(x_i - \mu_i)\sigma^{ip}\right] + \ldots\right\}$$

and show that the variable x_p in the expression for the conditional density has the form

$$\exp\left\{-\frac{1}{2}\sigma^{pp}\left(x_p - \mu_p + \sum_{i=1}^{p-1}(x_i - \mu_i)\sigma^{ip}/\sigma^{pp}\right)^2\right\}.$$

1.64. An algorithm to simulate a random variable with Poisson's distribution $\Pi(\lambda)$ is based on the following fact (prove this!). Suppose that U_i, $i = 0, 1, 2, \ldots$, are independent random variables uniformly distributed as $R(0, 1)$ and $\xi = \max\left\{k\colon \prod_{i=1}^{k} U_i \geqslant e^{-\lambda}\right\}$, then $\mathscr{L}(\xi) = \Pi(\lambda)$.

Hint. Using the relation $\mathscr{L}\left(-\sum_{i=1}^{k}\ln U_i\right) = \Gamma(1, k)$, calculate the probability of the event $\{\xi = k\} = \left\{-\sum_{i=1}^{k}\ln U_i \leqslant \lambda;\ -\sum_{i=1}^{k+1}\ln U_i > \lambda\right\}$.

1.65. Let a bounded distribution density $f(x)$, $c = \max\limits_{a \leqslant x \leqslant b} f(x)$, be given on the segment $[a, b]$. We define the random variable

$$\nu = \min\{i \geqslant 1\colon cU_{2i-1} \leqslant f(a + (b - a)U_{2i})\},$$

where $\{U_i\}$ are as in the previous problem.

Prove that the random variable $\xi = a + (b - a)U_{2\nu}$ has the distribution density $f(x)$.

Remark. This result gives a simulation technique for a distribution with an arbitrary density, which satisfies the indicated restrictions.

Estimation of Distribution Parameters

2.1. Suppose that we have a statistical model $\mathscr{F} = \{F\}$ for a scheme of repeated independent observations on a random variable ξ and $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample from $\mathscr{L}(\xi)$. Any random variable $T = T(\mathbf{X})$ which is only a function of the sample is called a *statistic*. Given a sample $\mathbf{X}$, we often have to estimate the true value of an unknown theoretical characteristic $g = g(F)$, i.e., to construct a statistic $T(\mathbf{X})$ which can be used as a reasonable approximation of the true characteristic g. In this case the statistic $T(\mathbf{X})$ is called an *estimator* for (of) g. Various estimators are used to estimate g, and we may compare their quality by the *measure of accuracy* (the degree of closeness to the true value of the estimated characteristic). If we are given a class of estimators $\mathscr{T}_g$ and the measure of accuracy is chosen, then the estimator which optimizes this measure is called an *optimum estimator* (in the class $\mathscr{T}_g$).

The most popular measure of accuracy is a *standard* (*mean square*) *error* $\mathbf{E}(T(\mathbf{X}) - g)^2$. This measure brings about a respective optimality test, i.e., a *minimal mean-square-error test*. We often restrict ourselves to the class $\mathscr{T}_g$ of *unbiased estimators*, viz., $T = T(\mathbf{X}) \in \mathscr{T}_g \Leftrightarrow \mathbf{E}T = g$ $\forall F \in \mathscr{F}$. For the unbiased estimators $\mathbf{E}(T - g)^2 = \mathbf{D}T$, i.e., their variance is a measure of their accuracy, and a *minimal variance test* is in this case the optimality test. If a model $\mathscr{F}$ is parametric, i.e., $\mathscr{F} = \{F(x; \theta), \theta \in \Theta\}$, any theoretical characteristic is a function of the parameter θ. Then we are dealing with the estimation of parametric functions denoted $\tau(\theta)$. The statistic $T = T(\mathbf{X})$ is an unbiased estimator for $\tau(\theta)$ if the relation $\mathbf{E}_\theta T = \tau(\theta)$ $\forall \theta \in \Theta$ is true. A statistic T^* for which $\mathbf{D}_\theta T^* \leqslant \mathbf{D}T$ $\forall T \in \mathscr{T}_\tau$ and $\forall \theta \in \Theta$ is an optimum estimator in the class $\mathscr{T}_\tau$ of the unbiased estimators for the function $\tau = \tau(\theta)$. We sometimes use τ^* to denote T^* in order to stress that it is related to the function $\tau(\theta)$. The optimum estimator T^* (for a given model $\mathscr{F}$ and a given parametric function $\tau(\theta)$) does not always exist, but when it does exist, it is unique [7, p. 55]. It is important that the opti-

mality is *linear*, i.e., if T_j^* is an optimum estimator for $\tau_j = \tau_j(\theta)$, $j = 1, 2, \ldots$, then the statistic $\sum_j c_j T_j^*$ is an optimum estimator for a linear combination $\sum_j c_j \tau_j$ [7, p. 58].

Consistency is necessary for any estimation rule. This means that as the size n of a sample grows, the estimator must converge in probability to the estimated characteristic, whatever the true distribution of the observations. Thus, consistency is an asymptotic property of estimators (in contrast to unbiasedness and optimality). When we want to stress that the statistics we are studying depend on sample size, we label them with the subscript n. When investigating the rule for consistency, we use the following simple test [7, p. 91]. If $\mathbf{E}_\theta T_n = \tau(\theta) + \varepsilon_n$, $\mathbf{D}_\theta T_n = \delta_n$, and $\varepsilon_n = \varepsilon_n(\theta) \to 0$ (i.e., T_n is an *asymptotically unbiased estimator* for $\tau(\theta)$) and $\delta_n = \delta_n(\theta) \to 0$ as $n \to \infty$ for all $\theta \in \Theta$, then T_n is a *consistent estimator* for $\tau(\theta)$.

2.2. We now consider the general tests for existence of optimum estimators and the ways to construct them in the framework of the general parametric model $\mathscr{F} = \{F(x; \theta), \theta \in \Theta\}$. Let $f(x; \theta)$ be the distribution density of the observable random variable ξ (or the probability of the event $\{\xi = x\}$ in a discrete case), and let $\mathbf{x} = (x_1, \ldots, x_n)$ be a realization of the sample $\mathbf{X} = (X_1, \ldots, X_n)$. For fixed $\mathbf{x} \in \mathscr{X}$ the function $L(\mathbf{x}; \theta) = f(x_1; \theta) \ldots f(x_n; \theta)$ of the parameter $\theta \in \Theta$ is called a *likelihood function.* We will assume that $L(\mathbf{x}; \theta) > 0$ for all $\mathbf{x} \in \mathscr{X}$ and $\theta \in \Theta$ and it is differentiable with respect to θ. Moreover, the following rule for changing the order of differentiation and integration (when θ is a scalar parameter) is valid: *for any statistic T (in particular, for $T =$* const), *we have*

$$\frac{\partial}{\partial\theta} \int T(\mathbf{x}) L(\mathbf{x}; \theta)\, d\mathbf{x} = \int T(\mathbf{x}) \frac{\partial}{\partial\theta} L(\mathbf{x}; \theta)\, d\mathbf{x}, \quad d\mathbf{x} = dx_1 \ldots dx_n.$$

(Integration is carried out over the entire sample space $\mathscr{X}$, the integrals are assumed to be absolutely convergent for all $\theta \in \Theta$. For discrete models integration is replaced by summation.) Finally, we introduce the random variable

$$U(\mathbf{X}; \theta) \equiv \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial \ln f(X_i; \theta)}{\partial\theta},$$

which is called the *sample contribution*, and we will assume that

$$0 < \mathbf{E}_\theta U^2(\mathbf{X}; \theta) < \infty \ \forall\theta \in \Theta.$$

Models for which all these conditions are met are called *regular.*

For a regular model $\mathbf{E}_\theta U(\mathbf{X};\ \theta) = 0 \ \ \forall\theta \in \Theta$ the function $i_n(\theta) \equiv \mathbf{D}_\theta U(\mathbf{X};\ \theta) = \mathbf{E}_\theta U^2(\mathbf{X};\ \theta)$, which is called *(Fisher's) information function*, is defined. The quantity

$$i(\theta) \equiv i_1(\theta) = \mathbf{E}_\theta \left(\frac{\partial \ln f(X_1;\ \theta)}{\partial\theta}\right)^2 = -\mathbf{E}_\theta \left(\frac{\partial^2 \ln f(X_1;\ \theta)}{\partial\theta^2}\right)$$

is also called the *amount of (Fisher's) information contained in one observation* (the latter expression is used when the function $f(x;\ \theta)$ is twice differentiable with respect to θ). In the case of repeated independent observations $i_n(\theta) = ni(\theta)$.

The introduced notions can be generalized to the case of a vector parameter $\boldsymbol{\theta} = (\theta_1,\ \ldots,\ \theta_r)$. Then the random vector

$$\mathbf{U} = (U_1(\mathbf{X};\ \boldsymbol{\theta}),\ \ldots,\ U_r(\mathbf{X};\ \boldsymbol{\theta})),$$

where

$$U_j(\mathbf{X};\ \boldsymbol{\theta}) = \frac{\partial}{\partial\theta_j} \ln L(\mathbf{X};\ \boldsymbol{\theta}),\quad j = 1,\ \ldots,\ r,$$

is a sample contribution, and the *information matrix* $\mathbf{I}_n = \mathbf{I}_n(\boldsymbol{\theta}) \equiv \mathbf{D}_{\boldsymbol{\theta}}(\mathbf{U}) = \mathbf{E}_{\boldsymbol{\theta}}(\mathbf{U}\mathbf{U}')$ *of the sample* is an analogue of the information function. The information matrix $\mathbf{I}_1 = \mathbf{I} = \|g_{ij}\|_1^r$ can be calculated by the formulas

$$g_{ij} = g_{ij}(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}} \left(\frac{\partial \ln f(X_1;\ \boldsymbol{\theta})}{\partial\theta_i}\,\frac{\partial \ln f(X_1;\ \boldsymbol{\theta})}{\partial\theta_j}\right)$$

$$= -\mathbf{E}_{\boldsymbol{\theta}} \left(\frac{\partial^2 \ln f(X_1;\ \boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right),$$

the last equation being true if the functions $f(x;\ \boldsymbol{\theta})$ are twice differentiable.

For repeated independent observations we have $\mathbf{I}_n(\boldsymbol{\theta}) = n\mathbf{I}(\boldsymbol{\theta})$. In this case the definition of a regular model implies that the matrix $\mathbf{I}(\boldsymbol{\theta})$ is non-singular for all $\boldsymbol{\theta} \in \Theta$.

We can find lower bounds for variances of unbiased estimators for a given differentiable parametric function $\tau(\theta)$ in a regular model. Indeed, *for any estimator* $T = T(\mathbf{X}) \in \mathscr{T}_\tau$ *and all* $\theta \in \Theta$ *the inequality*

$$\mathbf{D}_\theta T \geqslant \frac{[\tau'(\theta)]^2}{ni(\theta)}$$

holds for a scalar parameter θ, *and the inequality*

$$\mathbf{D}_\theta T \geqslant \mathbf{b}'(\theta)\mathbf{I}_n^{-1}(\theta)\mathbf{b}(\theta), \quad \mathbf{b}(\theta) = \left(\frac{\partial\tau(\theta)}{\partial\theta_1}, \ldots, \frac{\partial\tau(\theta)}{\partial\theta_r}\right)$$

holds for a vector parameter $\theta = (\theta_1, \ldots, \theta_r)$. This is the *Cramér-Rao inequality*. The estimator $T^* \in \mathscr{T}_\tau$ for which the indicated lower bound is reached is said to be *efficient*. If an efficient estimator exists, it is, consequently, optimal (in the class $\mathscr{T}_\tau$) and unique. The representation [7, p. 61]

$$T(\mathbf{X}) - \tau(\theta) = a(\theta)U(\mathbf{X};\, \theta) \quad \text{if} \quad \theta \quad \text{is a scalar,}$$
$$T(\mathbf{X}) - \tau(\theta) = \mathbf{a}'(\theta)\mathbf{U}(\mathbf{X};\, \theta) \quad \text{if} \quad \theta \quad \text{is a vector}$$

is a test for efficiency. Here $a(\theta)$ ($\mathbf{a}'(\theta)$) is a function (vector function) of θ.

In a given model $\mathscr{F}$ an efficient estimator can only exist for one parametric function $\tau(\theta)$ (up to the transformation $a\tau(\theta) + b$, where a and b are constants).

If there is no efficient estimator, then we use the *Bhattacharyya test* [7, p. 64] to find an optimum estimator $T^* = \tau^*$ (in the class of unbiased estimators $\mathscr{T}_\tau$), i.e., taking into account the higher-order derivatives of the likelihood function $L = L(\mathbf{X};\, \theta)$, we choose a linear combination of them in order to obtain a representation of the form

$$T - \tau = \frac{1}{L}\left(\sum_i a_i \frac{\partial L}{\partial\theta_i} + \sum_{i,j} a_{ij}\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} + \ldots + \sum_{i_1, \ldots, i_s} a_{i_1\ldots i_s}\frac{\partial^s L}{\partial\theta_{i_1}\ldots\partial\theta_{i_s}}\right).$$

We successively put here $s = 2, 3, \ldots$. If we manage to do so for some $s \geqslant 2$ and the coefficients $a_. = a_.(\theta)$, then the statistic $T = T(\mathbf{X})$ is an optimum estimator for the function $\tau = \tau(\theta)$.

2.3. The most effective way to construct optimum estimators is to use sufficient statistics. A statistic $T = T(\mathbf{X})$ (generally a vector statistic) is said to be *sufficient* for the model $\mathscr{F} = \{F(x;\, \theta), \theta \in \Theta\}$ (or for the parameter θ) if the conditional density (probability in the discrete case) $L(\mathbf{x}\,|\,t;\, \theta)$ of a random vector (sample) $\mathbf{X} = (X_1, \ldots, X_n)$ is independent of the parameter θ under the condition $T(\mathbf{X}) = t$. We may use an equivalent definition, i.e., *for any event* $A \subset \mathscr{X}$ *the conditional probability* $\mathbf{P}_\theta(\mathbf{X} \in A\,|\,T(\mathbf{X}) = t)$ *is independent of* θ. This property of

the statistic T implies that it covers all the information about the parameter θ contained in a sample. Indeed, the probability of any event which can occur at a fixed T is independent of θ and, hence, it has no additional information about θ. The sample $\mathbf{X}$ is obviously a sufficient statistic, but we usually seek a smallest-dimensional sufficient statistic which represents the original data in the most compact form, i.e., we seek a *minimal sufficient statistic*. A minimal sufficient statistic is a function of any other sufficient statistics. We use the *factorization test* [7, p. 70] to construct sufficient statistics, i.e., *a statistic $T(\mathbf{X})$ is sufficient for the parameter θ if and only if the likelihood function can be represented as*

$$L(\mathbf{x};\ \theta) = g(T(\mathbf{x});\ \theta)h(\mathbf{x}),$$

where g and h are non-negative functions, and h is independent of θ. If T is a sufficient statistic, then any other function which is in a one-to-one correspondence with T is also a sufficient statistic.

The *Rao-Blackwell-Kolmogorov theorem* [7, p. 72] defines the role of sufficient statistics in estimation theory. The theorem states that *for any unbiased estimator T_1 of a given function $\tau(\theta)$ we can construct a new unbiased estimator $T^* \equiv \mathbf{E}_\theta(T_1 | T)$ which depends on the sufficient statistic T and obeys the inequality $\mathbf{D}_\theta T^* \leqslant \mathbf{D}_\theta T_1$*. Consequently, an optimum estimator should be sought among the functions of a sufficient statistic.

We use an important property of completeness of a sufficient statistic in order to find an explicit form of optimum estimators. The statistic T is said to be (boundedly) *complete* if for any (bounded) function $\varphi(T)$ the equation $\mathbf{E}_\theta \varphi(T) = 0\ \forall\theta$ implies that $\varphi(t) \equiv 0$ on the domain of T.

If a complete sufficient statistic exists, then every function of it is an optimum estimator of its mean. Consequently, when estimating a given parametric function $\tau(\theta)$, we find an optimum unbiased estimator τ^*, which is a function $\tau^* = H(T)$ of a complete sufficient statistic T and satisfies the *unbiasedness equation* $\mathbf{E}_\theta H(T) = \tau(\theta)$. This equation either has a unique solution or has no solution. In the latter case the class $\mathcal{T}_\tau$ of unbiased estimators $\tau(\theta)$ is empty.

Many models in mathematical statistics belong to an *r-parametric exponential family*, i.e., for them the function $f(x;\ \theta)$, $\theta = (\theta_1, \ldots, \theta_r) \in \Theta \subset R^r$ can be represented in the form

$$f(x;\ \theta) = \exp\left\{ \sum_{j=1}^{r} \theta_j B_j(x) + C(\theta) + D(x) \right\}$$

(or it can be reduced to this form by a change of the parameters).

Then $T = (T_1, \ldots, T_r)$, $T_j = T_j(\mathbf{X}) = \sum_{i=1}^{n} B_j(X_i)$, $j = 1, \ldots, r$, is a minimal sufficient statistic and it is complete if $\dim \Theta = r$.

2.4. The *maximum likelihood method* is universal among the estimation methods for unknown parameters. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$, we seek for a *maximum likelihood estimate* (m.l.e.) $\hat{\theta}_n$ for the parameter θ, which is a point of parametric set Θ where the maximum likelihood function $L(\mathbf{x}; \theta)$ attains its maximum for every $\mathbf{X} = \mathbf{x}$, i.e., $L(\mathbf{x}; \hat{\theta}_n) \geqslant L(\mathbf{x}; \theta)\ \forall \theta$, or $L(\mathbf{x}; \hat{\theta}_n) = \sup_{\theta \in \Theta} L(\mathbf{x}; \theta)$. If for any $\mathbf{x} \in \mathscr{X}$ the function $L(\mathbf{x}; \theta)$ attains its maximum at an internal point of Θ and $L(\mathbf{x}; \theta)$ is differentiable with respect to θ, then the m.l.e. $\hat{\theta}_n$ meets the *likelihood equation* $\dfrac{\partial \ln L(x; \theta)}{\partial \theta} = 0$ $\left(\text{or } \dfrac{\partial \ln L(x; \theta)}{\partial \theta_j} = 0,\ j = 1, \ldots, r, \text{ if } \theta = (\theta_1, \ldots, \theta_r)\right)$.

For a parametric function $\tau(\theta)$, the m.l.e. is $\hat{\tau}_n = \tau(\hat{\theta}_n)$. This is the *invariance* of maximum likelihood estimates.

When a likelihood equation cannot be solved exactly, we turn to approximate methods of solution. One of them is a recurrent *accumulation method* (due to Fisher), according to which the $(k + 1)$th approximation for a m.l.e. is computed by the formula

$$\theta_{k+1} = \theta_k + U(\mathbf{x}; \theta_k)/ni(\theta_k), \quad k = 0, 1, 2, \ldots.$$

Here we choose an easily computed consistent estimate for θ as a first approximation θ_0.

In the case of regular models the maximum likelihood estimates have important asymptotic properties. Indeed [7, pp. 92-95], if $\hat{\theta}_n$ exists, is unique, and lies inside Θ, then it is a consistent estimate for θ and its distribution is asymptotically normal, viz.,

$$\mathscr{L}_\theta(\sqrt{n}(\hat{\theta}_n - \theta)) \to \mathscr{N}(0, \mathbf{I}^{-1}(\theta))$$

as $n \to \infty$. If we additionally assume that the function $f(x; \theta)$ is three times differentiable with respect to θ and $\left|\dfrac{\partial^2 f(x; \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k}\right| \leqslant M(x)$, where the function $M(x)$ is independent of θ and integrable, then $\mathbf{E}_\theta M(\xi) < \infty$. If the elements of the matrix $\mathbf{I}(\theta)$ are continuous in θ, then

$$\mathscr{L}_\theta(\sqrt{n}(\hat{\theta}_n - \theta)) \sim \mathscr{N}(0, \mathbf{I}^{-1}(\hat{\theta}_n)).$$

Moreover, if $\tau(\theta)$ is a continuously differentiable function and

$\hat{\tau}_n = \tau(\hat{\theta}_n)$ is its m.l.e., then

$$\mathscr{L}_\theta(\sqrt{n}(\hat{\tau}_n - \tau(\theta))) \to \mathscr{N}(0, \sigma_\tau^2(\theta))$$

as $n \to \infty$, and

$$\mathscr{L}_\theta\left(\sqrt{n}\,\frac{\hat{\tau}_n - \tau(\theta)}{\sigma_\tau(\hat{\theta}_n)}\right) \to \mathscr{N}(0, 1),$$

where $\sigma_\tau^2(\theta) = \mathbf{b}'(\theta)\mathbf{I}^{-1}(\theta)\mathbf{b}(\theta)$, $\mathbf{b}(\theta) = \left(\frac{\partial\tau(\theta)}{\partial\theta_1}, \ldots, \frac{\partial\tau(\theta)}{\partial\theta_r}\right)$. The quantity $\sigma_\tau^2(\theta)/n$ is called the *asymptotic variance* of the statistic $\hat{\tau}_n$ and coincides with the Cramér-Rao bound for the variances of the unbiased estimators for the function $\tau(\theta)$. This property of maximum likelihood estimates is called the *asymptotic efficiency*.

If we have another consistent and asymptotically normal estimator T_n for the function $\tau(\theta)$, viz., $\mathscr{L}_\theta(\sqrt{n}(T_n - \tau(\theta))) \to \mathscr{N}(0, \sigma_T^2(\theta))$ as $n \to \infty$, then its "quality" can be measured by the quantity eff $(T_n; \theta) \equiv \sigma_\tau^2(\theta)/\sigma_T^2(\theta)$, which is called the *asymptotic efficiency* of the estimator T_n. The estimator is the better (the more exact) asymptotically, the greater its asymptotic efficiency. For m.l.e.'s this quantity is equal to unity.

2.5. Until now we have considered point estimation of the unknown distribution parameters though mathematical statistics also deals with *confidence interval estimation* or (for vector parameters) with *estimation by confidence sets.* Let θ be a scalar. In interval estimation we seek two statistics $T_i = T_i(\mathbf{X})$, $i = 1, 2$, such that $T_1 < T_2$. For these statistics the condition

$$\mathbf{P}_\theta(T_1(\mathbf{X}) < \theta < T_2(\mathbf{X})) \geqslant \gamma \quad \forall\theta \in \Theta \tag{$*$}$$

must hold at a *confidence level* $\gamma \in (0, 1)$. This (random) interval $(T_1, T_2) \subset \Theta$ is called a γ-*confidence interval* for θ. Its length $T_2 - T_1$ characterizes the accuracy of the unknown parameter localization, while the γ-confidence level characterizes its "reliability", i.e., the probability that the assertion $\theta \in (T_1, T_2)$ is erroneous does not exceed $1 - \gamma$. In practice γ is taken to be close to 1 ($\gamma = 0.95, 0.99$, etc.), and then the shortest (for a given class) interval is constructed for the chosen γ.

We sometimes use *one-sided confidence intervals* (*upper*, of the form $\theta < T_2(\mathbf{X})$, and *lower*, of the form $T_1(\mathbf{X}) < \theta$), which are defined by the conditions similar to ($*$) but without the second limit.

In the case of a vector parameter the confidence interval for a separate component (for example, θ_1) is chosen in a similar way, viz.,

$$\mathbf{P}_\theta(T_1(\mathbf{X}) < \theta_1 < T_2(\mathbf{X})) \geqslant \gamma \quad \forall\theta \in \Theta,$$

as well as the confidence interval for a parametric function $\tau(\theta)$

$$\mathbf{P}_\theta(T_1(\mathbf{X}) < \tau(\theta) < T_2(\mathbf{X})) \geqslant \gamma \quad \forall\theta \in \Theta.$$

A *γ-confidence region* for a vector parameter $\theta = (\theta_1, \ldots, \theta_r)$ is a random subset $\mathscr{S}_\gamma(\mathbf{X}) \subset \Theta$ for which

$$\mathbf{P}_\theta(\theta \in \mathscr{S}_\gamma(\mathbf{X})) \geqslant \gamma \quad \forall\theta \in \Theta.$$

This subset is constructed using a statistic $T(\mathbf{X})$ whose distribution is known.

If we are estimating a scalar parameter θ and know that there exists a random variable $G(\mathbf{X};\ \theta)$ which depends on the observations $\mathbf{X} = (X_1, \ldots, X_n)$ and on the estimated parameter, such that (1) the distribution of $G(\mathbf{X};\ \theta)$ is independent of θ and (2) for every $\mathbf{x} \in \mathscr{X}$ the function $G(\mathbf{x};\ \theta)$ is continuous and strictly monotone in θ (in this case $G(\mathbf{X};\ \theta)$ is called a *central statistic*), then the γ-confidence interval for θ is constructed in the following way. We define the numbers $g_1 < g_2$ from the condition $\mathbf{P}_\theta(g_1 < G(\mathbf{X};\ \theta) < g_2) = \gamma$ and solve the equations $G(\mathbf{X};\ \theta) = g_1, g_2$ with respect to θ. We use $T_i = T_i(\mathbf{X})$, $i = 1, 2$, $T_1 < T_2$, to denote the solutions and find the required interval (T_1, T_2). The technique based on central statistics can be applied to estimate the components of a parametric vector $\theta = (\theta_1, \ldots, \theta_r)$ and scalar parametric functions $\tau = \tau(\theta)$.

If we have some point estimator $T = T(\mathbf{X})$ for the parameter θ, and its distribution function $F(t;\ \theta)$ is continuous and monotone in θ, then, having found from the equations (with respect to θ)

$$F(T;\ \theta) = (1 - \gamma)/2, \quad 1 - F(T - 0;\ \theta) = (1 - \gamma)/2$$

two random numbers $T_i = T_i(\mathbf{X})$, $i = 1, 2$, $T_1 < T_2$, we find the *central γ-confidence interval* (T_1, T_2) for θ.

It is sometimes possible to construct *approximate confidence intervals* for large samples using maximum likelihood estimates. Thus, if $\tau(\theta)$, $\theta = (\theta_1, \ldots, \theta_r)$, is a continuously differentiable function and $\hat{\tau}_n = \tau(\hat{\theta}_n)$ is its maximum likelihood estimate, then, in the case of a regular model, the interval $(\hat{\tau}_n \pm c_\gamma \sigma_\tau(\hat{\theta}_n)/\sqrt{n})$ is an *asymptotic γ-confidence interval* for $\tau(\theta)$, where $\sigma_\tau^2(\theta) = \mathbf{b}'(\theta)\mathbf{I}^{-1}(\theta)\mathbf{b}(\theta)$, $\mathbf{b}(\theta) = \left(\dfrac{\partial\tau(\theta)}{\partial\theta_1}, \ \ldots, \ \dfrac{\partial\tau(\theta)}{\partial\theta_r}\right)$, $c_\gamma = \Phi^{-1}\left(\dfrac{1 + \gamma}{2}\right)$. Specifically, $(\hat{\theta}_n \pm c_\gamma/\sqrt{ni(\hat{\theta}_n)})$ is an asymptotic γ-confidence interval for a scalar parameter θ. Such intervals are asymptotically shortest and are based on the standard normal approximation

$$\mathscr{L}_\theta(\hat{\tau}_n) \sim \mathscr{N}(\tau(\theta),\ \sigma_\tau^2(\hat{\theta}_n)/n)$$

for a m.l.e.

Problems

Estimators and Their General Properties

2.1. Show that the following statistics are unbiased and consistent:

(a) $T_n(\mathbf{X}) = F_n(x)$ as the point estimators for the theoretical distribution function $F(x)$ at a given point x;

(b) $T_n(\mathbf{X}) = A_{nk}$ as the estimators for the theoretical moment $\alpha_k = \mathbf{E}\xi^k$;

(c) $T_n(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \alpha_1)^2$ as the estimators for the variance $\mu_2 = \mathbf{D}\xi$ when the mean $\alpha_1 = \mathbf{E}\xi$ is known;

(d) $T_n(\mathbf{X}) = \frac{n}{n-1}S^2 \equiv S'^2$ as the estimators for μ_2 in the general case. Check whether S^2 is a consistent estimator for μ_2.

Hint. Use Problem 1.27 and the Chebyshev inequality (assume that the respective theoretical moments do exist).

2.2. In what cases is the statistic $T_n(\mathbf{X}) = \sqrt{A_{n2}/2}$ a consistent estimator for the theoretical mean α_1?

2.3. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $\mathscr{L}(\xi)$, construct an unbiased estimator for its characteristic function.

Hint. Consider an empirical distribution function.

2.4. Let $\mathbf{X} = ((X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2}))$ be a sample from a distribution of a two-dimensional random variable $\xi = (\xi_1, \xi_2)$. Prove that the statistic $T(\mathbf{X}) = \frac{n}{n-1}S_{12}$ is an unbiased estimator for $\mu_{11} = \operatorname{cov}(\xi_1, \xi_2)$, where S_{12} is the sample covariance (see the solution to Problem 1.38).

Hint. Consider the random variable $\xi_1 + \xi_2$ and use the solution to Problem 2.1 (d).

2.5. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $Bi(1, \theta)$. Describe the class of parametric functions $\tau(\theta)$ for which the unbiased estimators $T(\mathbf{X})$ exist. Show that the functions $\tau(\theta) = 1/\theta^a$ for $a > 0$ and $\tau(\theta) = \theta^b$ for $b > n$ do not belong to this class.

2.6*. Given the results of n trials, estimate the unknown probability of success θ in the Bernoulli trials $Bi(1, \theta)$. Use r_n to denote the number of successes in these trials and consider the class of estimators $T = \frac{r_n + \alpha}{n + \beta}$. Compute the standard error of the estimator T and compare it with the error of an "ordinary" estimator r_n/n.

2.7. Suppose that $\mathscr{L}(\xi) = Bi(k, \theta)$ and $n = 1$. Consider the functions of the form $\tau_{rs}(\theta) = \theta^r(1-\theta)^s$ for integer $r, s \geqslant 0$. Show that

the unbiased estimator for $\tau_{rs}(\theta)$ exists if and only if $r + s \leqslant k$ and then it has the form

$$T(X) = (X)_r(k - X)_s/(k)_{r+s},$$

where $(a)_r = a(a-1)\dots(a-r+1)$, $r \geqslant 1$, $(a)_0 = 1$.

2.8. Let $\mathbf{X} = (X_1, \dots, X_n)$ be a sample from the distribution $Bi(k, \theta)$ and $T = X_1 + \dots + X_n$. Describe the class of parametric functions $\tau(\theta)$ for which there exist unbiased estimators of the form $H(T)$. Construct such an unbiased estimator for $\tau_j(\theta) = \theta^j$.

Hint. Use the reproducibility of the distribution $Bi(k, \theta)$ (see Problem 1.39 (3) and Problem 1.52 (b)).

2.9. Suppose that $\mathscr{L}(\xi) = \Pi(\theta)$ and $n = 1$. Check whether $T(X) = (X)_j$ is an unbiased estimator for $\tau(\theta) = \theta^j$, $j = 1, 2, \dots$, and show that there are no unbiased estimators for the functions $\tau(\theta) = \theta^{-a}$, where $a > 0$. Construct an unbiased estimator for $(1 + \theta)^{-1}$.

2.10. Given one observation on a discrete random variable ξ with the distribution $f(x; \theta) = e^{-\theta}\dfrac{\theta^x}{x!}\Big/(1 - e^{-\theta})$, $x = 1, 2, \dots$ (*Poisson's distribution truncated at zero*), estimate the function $\tau(\theta) = 1 - e^{-\theta}$. Show that only one and practically useless estimator exists for it.

2.11. Let $\mathscr{L}(\xi) = \overline{Bi}(r, \theta)$ and $n = 1$. Construct an unbiased estimator for the function $\tau(\theta) = \theta^s$ ($s \geqslant 1$ is an integer) and make sure that for $r = 1$ this estimator is practically useless.

Hint. Use the formula $(1 - \theta)^{-r} = \sum\limits_{j=0}^{\infty} C^j_{r+j-1}\theta^j$.

2.12. Show that $T(X) = \sum\limits_{i=1}^{X} \dfrac{1}{i}$ is the only unbiased estimator for the function $\tau(\theta) = \ln(1 - \theta)$ in the model $\overline{Bi}(1, \theta)$ for $n = 1$.

2.13. Make sure that $T^* = \overline{X}^2 - \dfrac{\sigma^2}{n}$ is an unbiased estimator for the function $\tau(\theta) = \theta^2$ in the model $\mathscr{N}(\theta, \sigma^2)$.

2.14. Given a sample of size n in the model $\mathscr{N}(\mu, \theta^2)$, estimate $\tau(\theta) = \theta^2$. Show that the sample variance S^2 has a smaller standard error than the unbiased estimator $\tau^* = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \mu)^2$. Which of the unbiased estimators (τ^* or S'^2) is more exact? (See Problem 2.1 (d).)

2.15. Prove that $T_n(\mathbf{X}) = \sqrt{\dfrac{\pi}{2}}\dfrac{1}{n}\sum\limits_{i=1}^{n}|X_i - \mu|$ is an unbiased and consistent estimator for the parameter θ in the model $\mathscr{N}(\mu, \theta^2)$.

2.16*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathcal{N}(\mu, \theta^2)$ and $T^2 = \sum_{i=1}^{n} (X_i - \mu)^2$. Prove that the statistic

$$\tau_k^* = \frac{\Gamma\left(\frac{n}{2}\right)}{2^{\frac{k}{2}}\Gamma\left(\frac{n+k}{2}\right)} T^k$$ is an unbiased estimator for the function $\tau_k(\theta) = \theta^k$ for any integer $k \geqslant 1$. Compare the estimator τ_1^* with that from the previous problem.

| *Hint.* Use the fact that $\mathcal{L}_\theta(T^2/\theta^2) = \chi^2(n)$.

2.17. Suppose that we are given a sample $\mathbf{X} = (X_1, X_2, X_3)$ from the distribution $\mathcal{N}(0, \theta^2)$ and $T = T(\mathbf{X}) = \sqrt{X_1^2 + X_2^2 + X_3^2}$. Consider a statistic $p_T(x) = \frac{1}{2T} I(|x| \leqslant T)$, where $I(\cdot)$ is an indicator, which is a function of the variable x and represents the uniform distribution density on the segment $[-T, T]$. Show that $p_T(x)$ is an unbiased estimator for the density of the original distribution $\mathcal{N}(0, \theta^2)$ for any x.

| *Hint.* Use the fact that $\mathcal{L}_\theta(T^2/\theta^2) = \chi^2(3)$.

2.18. Let us estimate an unknown variance θ_2^2 in the general normal model $\mathcal{N}(\theta_1, \theta_2^2)$. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the respective sample, and let S'^2 be the unbiased estimator for θ_2^2 (see Problem 2.1 (d)). Consider the class of estimators of the form $T_\lambda = \lambda S'^2$. Show that for $\frac{n-3}{n+1} < \lambda < 1$ the statistic T_λ has a smaller standard error than S'^2. For what integer k do the statistics $\frac{1}{n+k} \sum_{i=1}^{n} (X_i - \overline{X})^2$ belong to this subclass? Using the minimal mean-square-error test, find an optimum estimator in the class $\{T_\lambda\}$.

2.19*. (Continued from Problem 2.18.) Construct optimum estimators of the form $T_\lambda = \lambda S'^2$, which minimize the measure $\mathbf{E}_\theta(T_\lambda - \theta_2^2)^4$ and $\mathbf{E}_\theta|T_\lambda - \theta_2^2|$, respectively.

| *Hint.* Use Fisher's theorem and Problem 1.45.

2.20. Prove that the statistic

$$\tau_k^* = \left(\frac{n}{2}\right)^{\frac{k}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n+k-1}{2}\right)} S^k,$$

where S^2 is a sample variance, is an unbiased estimator for the function $\tau_k(\theta) = \theta_2^k$ and at integer $k \geqslant 1$ in the model of Problem 2.18. Consider the case of $n = 2$ and calculate the bias of the statistic $|X_1 - X_2|$ which is an estimator for θ_2.

| *Hint.* Take into account that $\mathscr{L}(nS^2/\theta_2^2) = \chi^2(n-1)$.

2.21. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\Gamma(\theta, \lambda)$ and $T = X_1 + \ldots + X_n$. Make sure that the statistic $\tau_a^* = \dfrac{\Gamma(\lambda n)}{\Gamma(\lambda n - a)} T^{-a}$ is an unbiased estimator for the function $\tau_a(\theta) = \theta^{-a}$ for any $a < \lambda n$.

| *Hint.* Use the reproducibility property of the gamma distribution (see Problem 1.39 (2)).

2.22*. The lifetime of electric bulbs is distributed as $\Gamma(\theta, 1)$. In order to estimate θ, we take a sample of n bulbs and observe the lifetimes of the first r burnt-out bulbs $X_{(1)} < X_{(2)} < \ldots < X_{(r)}$. Construct an optimum unbiased estimator of the form $T(\mathbf{X}) = \sum_{k=1}^{r} \lambda_k X_{(k)}$.

| *Hint.* Use the random variables $Y_r = \dfrac{n - r + 1}{\theta}(X_{(r)} - X_{(r-1)})$, $r = 1, \ldots, n$, $(X_{(0)} = 0)$, and Problem 1.34.

2.23. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $R(\theta, 2\theta)$, estimate the parameter θ. Consider a class of estimators of the form $T = T(\mathbf{X}) = \alpha X_{(n)} + \beta X_{(1)}$, $\alpha, \beta \geqslant 0$, and find an optimum unbiased estimator in it.

| *Hint.* Use Problem 1.36.

2.24. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$, estimate the parameter θ of the uniform distribution $R(0, \theta)$. Show that the statistics $T_1 = \dfrac{n+1}{n} X_{(n)}$ and $T_2 = (n+1)X_{(1)}$ are unbiased. Which of them is better?

| *Hint.* Use Problem 1.36. Show that T_2 is not a consistent estimator.

2.25. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $R(\theta_1, \theta_2)$. Prove that the statistics $T_1 = (X_{(1)} + X_{(n)})/2$ and $T_2 = \dfrac{n+1}{n-1}(X_{(n)} - X_{(1)})$ are unbiased and consistent estimators for the functions $\tau_1(\theta) = (\theta_1 + \theta_2)/2$ and $\tau_2(\theta) = \theta_2 - \theta_1$, respectively.

| *Hint.* Use Problem 1.36.

2.26. Prove that if $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample from Weibull's distribution $W(\theta, \alpha, b)$ with an unknown shift parameter θ, then the

statistic $T(\mathbf{X}) = X_{(1)} - b\Gamma\left(1 + \frac{1}{\alpha}\right)n^{-1/\alpha}$ is an unbiased and consistent estimator for the parameter θ.

Hint. Use the solution to Problem 1.37.

2.27. Show that for a *logistic distribution* with the density $f(x;\ \theta) = e^{-x+\theta}(1 + e^{-x+\theta})^{-2}$, $-\infty < x < \infty$, $\theta \in (-\infty,\ \infty)$, the sample mean $\overline{X}$ is an unbiased and consistent estimator for the parameter θ.

2.28. Show that the sample mean $\overline{X}$ is not a consistent estimator for the parameter θ in the Cauchy model $C(\theta)$.

Hint. Use the property of an arithmetic mean of Cauchy's distribution.

2.29. *Estimation of a polynomial distribution.* Let a random variable ξ have a finite number of values $a_1, \ldots, a_N$ with unknown probabilities $p_1, \ldots, p_N$, $p_1 + \ldots + p_N = 1$. In order to estimate the parameter $\theta = (p_1, \ldots, p_{N-1})$, where $p_N = 1 - p_1 - \ldots - p_{N-1}$, we carry out n independent observations on ξ. Let ν_r be the number of units in a sample, which are equal to a_r, $r = 1, \ldots, N$.

(a) Show that the statistics $T_r = \nu_r/n$, $r = 1, \ldots, N$, are unbiased and consistent estimators for the parameters $p_1, \ldots, p_N$, respectively.

(b) Describe the class of the parametric functions $\tau(\theta)$ for which the unbiased estimators of the form $H(T_1, \ldots, T_N)$ exist.

(c) Construct an unbiased and consistent estimator for the function $\tau(\theta) = \sum_{i=1}^{N} c_i p_i$.

Hint. Take into account that $\mathscr{L}(\nu_1, \ldots, \nu_N) = M(n; p_1, \ldots, p_N)$ and use the solution to Problem 1.52.

2.30. *Estimation by the method of moments.* Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathscr{L}(\xi) \in \{F(x;\ \theta),\ \theta = (\theta_1, \ldots, \theta_r) \in \Theta\}$ and suppose that the moments $\alpha_k(\theta) = \mathbf{E}_\theta \xi^k$, $k = 1, \ldots, r$, exist. Then, solving the equations $\alpha_k(\theta) = A_{nk}$, $k = 1, \ldots, r$, in $\theta_1, \ldots, \theta_r$, where $A_{nk} = A_{nk}(\mathbf{X})$ is the sampling moment of kth order, we find the estimates for the parameters using the method of moments.

Using the method of moments, find the estimates for the parameters of the gamma distribution $\Gamma(\theta_1, \theta_2)$ and make sure that they are consistent.

2.31. Using the method of moments, find the estimates for the parameters of a *bivariate Poisson distribution* given by the probabilities

$$\mathbf{P}_\theta(\xi = x) = \frac{1}{2}\left(e^{-\theta_1}\frac{\theta_1^x}{x!} + e^{-\theta_2}\frac{\theta_2^x}{x!}\right),$$

$$x = 0, 1, 2, \ldots, \theta = (\theta_1, \theta_2), \quad 0 < \theta_1 < \theta_2.$$

This distribution describes, for example, the number of collisions of gas molecules in a Wilson's chamber with the particles formed when a uranium nucleus is bombarded by neutrons.

Calculate the estimates for the data obtained in $n = 327$ observations on a random variable ξ (n_x denotes the number of observations where $\xi = x$), viz.,

x	0	1	2	3	4	5	6	7	8	9	10
n_x	28	47	81	67	53	24	13	8	3	2	1

2.32. Simulate samples of sizes $n = 10, 100, 1000$ from the uniform distribution $R(0, \theta)$ for $\theta = 1$ and estimate the parameter θ using the method of moments.

2.33*. *Sampling inspection.* Statistical control over the quality of products may be carried out in the following way. We choose for control n items from a batch of N items at random and without replacement. Every of the n items is tested. If the number k of defective items in the sample meets the condition $k \leqslant k_0$, where k_0 is a preassigned level ($k_0 < n$), then these items are replaced by effective ones and the batch is accepted. If $k > k_0$, all the N items are inspected and the defectives are replaced. We use D to denote the unknown number of defectives in the batch ($D = 0, 1, \ldots, N$), and the random variable ξ to denote the number of defectives found in this way. Then

$$\mathbf{P}_D(\xi = k) = f(k;\, D,\, n) \equiv C_D^k C_{N-D}^{n-k}/C_N^n, \quad k = 0, 1, \ldots, k_0,$$

$$\mathbf{P}_D(\xi = D) = \sum_{k=k_0+1}^{n} f(k;\, D,\, n) \text{ (for } D > k_0).$$

Let us estimate a given function $\tau(D)$ of the number of defectives in a batch. Prove that a unique statistic $T(\xi)$ always exists and is an unbiased estimator for $\tau(D)$, i.e., the function $T(k)$ is uniquely defined by the conditions

$$\mathbf{E}_D T(\xi) = \sum_{k=0}^{k_0} T(k) f(k;\, D,\, n) + T(D) \sum_{k=k_0+1}^{n} f(k;\, D,\, n) = \tau(D),$$

$$D = 0, 1, \ldots, N.$$

Consider the case $\tau(D) = D$.

Hint. Use the fact that for $D \leqslant k_0$ the hypergeometric probabilities $f(k;\ D,\ n)$ are zero for $k > k_0$.

2.34*. *Estimation of a finite population.* Suppose that we have a finite population $U = \{u_1, \ldots, u_N\}$ of N objects each of which is characterized by a quantity $x(u)$, $u \in U$. The values $x_i = x(u_i)$, $i = 1, \ldots, N$, are unknown, and we are to estimate their sum $T = T(\mathbf{x}) = \sum_{i=1}^{N} x_i$. We stipulate that every subset (sample) $s = (u_{i_1}, \ldots, u_{i_{n(s)}})$ of units of U can be observed with a probability $p(s)$. Thus, if $S = \{s\}$ is a set comprising all the samples, then $\sum_{s \in S} p(s) = 1$. In this case we say that a *sampling plan* $\mathscr{A} = (U, S, \mathbf{P})$ is defined. The statistics $\bar{e}(s, \mathbf{x})$ which only depend on $\mathbf{x}$ through the x_i for which $u_i \in s$ are chosen as the estimators for T (i.e., the estimator is a function of the chosen objects and their observed x-values).

A statistic

$$\bar{e}(s,\ \mathbf{x}) = \sum_{u \in s} \frac{x(u)}{\pi(u)}$$

is called a *Horvitz-Thompson estimator*, where $\pi(u) = \sum_{s:\ s \ni u} p(s)$ is the probability that the object u is included in the sample.

(a) Prove that $\bar{e}(s, \mathbf{x})$ is an unbiased estimator for $T(\mathbf{x})$, i.e.,

$$\sum_s p(s)\bar{e}(s,\ \mathbf{x}) = T(\mathbf{x}) \quad \forall \mathbf{x} \in R^N,$$

and that there are no other unbiased estimators of the form $\sum_{u \in s} a(u)x(u)$.

(b) Derive the formula for the variance of the Horvitz-Thompson estimator

$$\mathbf{D}\bar{e}(s,\ \mathbf{x}) = \sum_{u \neq v} x(u)x(v)\left(\frac{\pi(u,\ v)}{\pi(u)\pi(v)} - 1\right) + \sum_u x^2(u)\left(\frac{1}{\pi(u)} - 1\right),$$

where $\pi(u, v) = \sum_{s:\ s \ni u,\ v} p(s)$ is the probability that the objects u and v are included in the sample.

(c) Check whether the statistic

$$\Delta(s, \mathbf{x}) = \sum_{u \in s} \frac{x^2(u)}{\pi(u)} \left(\frac{1}{\pi(u)} - 1\right) + \sum_{\substack{u, v \in s \\ u \neq v}} \frac{x(u)x(v)}{\pi(u, v)} \left(\frac{\pi(u, v)}{\pi(u)\pi(v)} - 1\right)$$

is an unbiased estimator for $\mathbf{D}\overline{e}(s, \mathbf{x})$.

(d) Show that the mean and variance of a sample s of size $n(s)$ for a sampling plan $\mathscr{A} = (\mathscr{U}, S, \mathbf{P})$ can be expressed through the probabilities of inclusion $\pi(u)$ and $\pi(u, v)$ in the following way:

$$\mathbf{E}n(s) = \sum_u \pi(u),$$

$$\mathbf{D}n(s) = \sum_{u \neq v} (\pi(u, v) - \pi(u)\pi(v)) + \sum_u \pi(u)(1 - \pi(u)).$$

Hint. Introduce the indicator random variables $\gamma(u) = I(u \in s)$ and write $n(s)$ and $\overline{e}(s, \mathbf{x})$ in the form

$$n(s) = \sum_u \gamma(u), \quad \overline{e}(s, \mathbf{x}) = \sum_u \gamma(u)x(u)/\pi(u).$$

2.35*. (Continued from Problem 2.34.) Consider a sampling plan $\mathscr{A}^* = (U, S, \mathbf{P})$ which generates unrepeated equiprobable samples of size n. In this case the set S consists of $(N)_n$ ordered combinations of size n of various units of U and

$$p(s) = \frac{1}{(N)_n} \quad \forall s \in S.$$

(a) Show that in our case the Horvitz-Thompson estimator has the form

$$\overline{e}(s, \mathbf{x}) = \frac{N}{n} \sum_{u \in s} x(u).$$

(b) We use

$$\mu = \frac{T(\mathbf{x})}{N}, \quad \sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2$$

to denote the *mean* and *variance of the population* U, respectively.

Then $\frac{1}{N}\,\overline{e}(s, \mathbf{x}) = \overline{x}$ (the sample mean of the observed x-values) is an unbiased estimator for μ. Check whether $\mathbf{D}\overline{x} = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2$.

(c) Prove that the statistic

$$\hat{\sigma}^2(s, \mathbf{x}) = \frac{1}{n-1}\sum_{u \in s}(x(u) - \overline{x})^2$$

is an unbiased estimator for σ^2.

Remark. A stronger result is true, i.e., $\hat{\sigma}^2(s, \mathbf{x})$ is an optimum estimator for σ^2 in the class of all the unbiased quadratic estimators of the form

$$\sum_{u,\, v \in s} a(u, v)(x(u) - \overline{x})(x(v) - \overline{x}).$$

2.36*. *Estimation of the size of a finite population.* Suppose that we are given a finite population U with an unknown number N of elements. We draw from it a simple unrepeated sample of size m and make this operation n times (each time any of the C_N^m possible combinations of elements of U can be drawn with the same probability). We use $\mu_r = \mu_r(n, m, N)$ to denote the number of the elements each of which appeared exactly r times ($r = 1, 2, \ldots, n$). We will estimate a parametric function $\tau(N)$ using the sample $(\mu_1, \ldots, \mu_n)$.

Prove that in the class of linear statistics $\mathscr{L} = \left\{l = \sum_{r=1}^{n} l_r\mu_r\right\}$ an unbiased estimator for $\tau(N)$ only exists when $\tau(N)$ is a polynomial in $1/N$ of degree $k \leqslant n-1$. In this case if $\tau(N) = \sum_{j=1}^{k} c_j/N^j$, then the statistic

$$\hat{\tau} = \sum_{r=1}^{n}\left[\sum_{j=1}^{k} c_j \frac{(r)_{j+1}}{m^{j+1}(n)_{j+1}}\right]\mu_r$$

is the only unbiased estimator for $\tau(N)$.

Specifically, $\frac{1}{m^2 n(n-1)}\sum_{r=1}^{n} r(r-1)\mu_r$ is the only linear unbiased estimator for $\tau(N) = 1/N$.

Hint. Represent μ_r as the sum of indicators, i.e., $\mu_r = \xi_1^{(r)} + \ldots + \xi_N^{(r)}$, where $\xi_i^{(r)} = 1$, if the ith element of U appeared r times, and $\xi_i^{(r)} = 0$ otherwise ($i = 1, \ldots, N$).

2.37*. (Continued from Problem 2.36.) Let $\eta = \mu_1 + \ldots + \mu_n$ be the total number of the observed elements, and let $\mathscr{H}$ be a class of statistics $H(\eta)$.

Prove that (a) if $N \leqslant mn$, then the statistic

$$\tau^* = \sum_{j=m}^{\eta} (-1)^{\eta-j} C_\eta^j (C_j^m)^n \tau(j) \Big/ \sum_{j=0}^{\eta} (-1)^{\eta-j} C_\eta^j (C_j^m)^n$$

is an unbiased estimator for any function $\tau(N)$ in the class $\mathscr{H}$;

(b) if N is a priori any integer ($N \geqslant m$), then this statistic is an unbiased estimator for the function $\tau(N)$ under the additional condition $\tau(N) = f(N)(C_N^m)^{-n}$, where $f(N)$ is a polynomial of a degree not exceeding mn and satisfying the conditions $f(0) = f(1) = \ldots = f(m-1) = 0$.

2.38. *The Monte Carlo method.* When seeking the values of various quantities (e.g., defined by equations or integrals), we often use a computational method based on their probabilistic interpretation and the realizations of random trials. This is the Monte Carlo method or the *method of statistical trials*. Depending on the nature of the calculated quantity a, we choose a random variable ξ so that $a = \mathbf{E}\xi$. We simulate a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $\mathscr{L}(\xi)$ and use the sample mean $\overline{X}$ to estimate a. Then (see Problem 1.61) we have

$$\mathbf{P}(\sqrt{n}|\overline{X} - a|/S' < c_\gamma) \to 2\Phi(c_\gamma) - 1 = \gamma$$

as $n \to \infty$. Thus, when n is large, the error in defining a by this method does not exceed $c_\gamma S'/\sqrt{n}$ with probability γ.

Suppose that we have to compute the value of the integral

$$a = \int_{v_r}\!\!\ldots\!\int f(t_1, \ldots, t_r)\, dt_1 \ldots dt_r,$$

where $v_r = \{(t_1, \ldots, t_r)\colon 0 \leqslant t_i \leqslant 1,\ i = 1, \ldots, r\}$. We obviously can take $\xi = f(\eta_1, \ldots, \eta_r)$, where $\eta_1, \ldots, \eta_r$ are independent random variables uniformly distributed on the segment $[0, 1]$, and simulate a sample $\mathbf{X}$ using the sequence (1.5).

Using the Monte Carlo method, estimate the integral $a = \int_0^1 e^x\, dx$, given 100 numbers from the sequence (1.5), and compare the resultant value of a^* with the exact value of a. For what δ will the relation $\mathbf{P}(|a - a^*| < \delta) \approx 0.99$ hold?

2.39. According to the Monte Carlo method, compute the value of the integral

$$p(r;\ \sigma_1,\ \sigma_2) = \frac{1}{2\pi} \iint\limits_{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 \leqslant r^2} \exp\left\{-\frac{1}{2}(x_1^2 + x_2^2)\right\} dx_1 dx_2$$

for $r = 3$, $\sigma_1 = 1$, $\sigma_2 = 2$, simulating a respective sample of size $n = 100$.

Hint. If ξ_1, ξ_2 are independent random variables and $\mathscr{L}(\xi_j) = \mathscr{N}(0,\ \sigma_j^2)$, $j = 1,\ 2$, then

$$p(r;\ \sigma_1,\ \sigma_2) = \mathbf{P}(\xi_1^2 + \xi_2^2 \leqslant r^2).$$

Now use Problem 1.61.

2.40*. *Random walk.* A particle starts at the moment $t = 0$ from the point k, $0 < k < N$, and wanders through the integer points of the interval $[0,\ N]$. If at the moment t the particle was at the point l, then at the moment $t + 1$ it will reach the point $l + 1$ with the probability p or the point $l - 1$ with the probability $q = 1 - p$, $1 \leqslant l \leqslant N - 1$. The walk is stopped at the points 0 or N, where the particle is annihilated [5, Chap. XIV].

(1) Find the probability π_{kN} that the particle will be annihilated at the point N.

(2) Compute $m_k = \mathbf{E}\tau_k$, where τ_k is the time before the annihilation.

(3) Simulate 100 realizations of the random walk just described for $N = 7$, $k = 3$, $p = 0.6$, $p = 0.5$ and find the estimates for π_{kN} and m_k.

Hints. (1) Write the finite-difference equation

$$f(k) = pf(k + 1) + qf(k - 1), \quad k = 1,\ \ldots,\ N - 1,$$
$$f(0) = 0, \quad f(N) = 1,$$

for $f(k) = \pi_{kN}$ and make sure that $\pi_{kN} = (1 - \lambda^k)/(1 - \lambda^N)$, $\lambda = q/p$, is its only solution if $p \neq q$, and $\pi_{kN} = k/N$ is its only solution if $p = q = 1/2$.

(2) Write the finite-difference equation

$$m_k = pm_{k+1} + qm_{k-1} + 1,$$
$$k = 1,\ \ldots,\ N - 1, \quad m_0 = m_N = 0,$$

for m_k and make sure that $m_k = \dfrac{k}{q - r}\dfrac{N}{q - p}\pi_{kN}$ is its only solution if $p \neq q$, and $m_k = k(N - k)$ is its only solution if $p = q = 1/2$.

(3) Act as in Problem 1.4.

Optimum Estimators

2.41*. Prove that an optimum unbiased estimator is always a symmetric function of observations.

Hint. If $T = T(\mathbf{X})$ is an unbiased estimator for $\tau(\theta)$, consider a symmetric statistic $T^* = \frac{1}{n!}\sum_{\pi} T(\pi\mathbf{X})$, where $\pi = \begin{pmatrix} 1 & \ldots & n \\ i_1 & \ldots & i_n \end{pmatrix}$ is a permutation of n elements, $\pi\mathbf{X} = (X_{i_1}, \ldots, X_{i_n})$, and summation is performed over all the $n!$ permutations. Show that $\mathbf{D}_\theta T^* \leqslant \mathbf{D}_\theta T$.

2.42. Prove the following properties of optimum estimators. If $T^* = T^*(\mathbf{X})$ is an optimum unbiased estimator for a function $\tau = \tau(\theta)$, then (1) an inequality $\operatorname{cov}_\theta(T^*, \psi) = 0\ \forall\theta$ holds for any statistic $\psi = \psi(\mathbf{X})$ with $\mathbf{E}_\theta\psi = 0\ \forall\theta \in \Theta$; (2) for any other unbiased estimator $T = T(\mathbf{X})$ $\operatorname{cov}_\theta(T^*, T) = \mathbf{D}_\theta T^*$.

Hint. In the first case consider unbiased estimators of the form $T_\lambda = T^* + \lambda\psi$, $\lambda \in R^1$. In the second case put $\psi = T^* - T$.

2.43. Check whether the amount of information $i(\theta)$ for the respective model has the form given in the following table:

Model	$\mathcal{N}(\theta, \sigma^2)$	$\mathcal{N}(\mu, \theta^2)$	$\Gamma(\theta, \lambda)$	$C(\theta)$	$Bi(k, \theta)$	$\Pi(\theta)$	$\overline{Bi}(r, \theta)$
$i(\theta)$	$1/\sigma^2$	$2/\theta^2$	λ/θ^2	$1/2$	$k/[\theta(1-\theta)]$	$1/\theta$	$r/[\theta(1-\theta)^2]$

2.44. Prove that the information matrix for the general normal model $\mathcal{N}(\theta_1, \theta_2^2)$ has the form

$$\mathbf{I}(\theta) = \begin{Vmatrix} 1/\theta_2^2 & 0 \\ 0 & 2/\theta_2^2 \end{Vmatrix}.$$

2.45*. Show that the information matrix for the model in Problem 2.29 has the form $\mathbf{I}(\theta) = \|g_{ij}(\theta)\|_1^{N-1}$, where $\theta = (p_1, \ldots, p_{N-1})$,

$$g_{ij}(\theta) = \begin{cases} 1/p_i + 1/p_N & \text{for } i = j, \\ 1/p_N & \text{for } i \neq j, \end{cases} \qquad p_N = 1 - p_1 - \ldots - p_{N-1}.$$

Calculate $\mathbf{I}^{-1}(\theta)$.

Hint. Write the probabilities $f(a_r;\ \theta) = \mathbf{P}_\theta(\xi = a_r) = p_r$, $r = 1, \ldots, N$, in the form

$$f(a_r;\ \theta) = \prod_{j=1}^{N} p_j^{\delta(a_r,\ a_j)} = (1 - p_1 - \ldots - p_{N-1})^{\delta(a_r,\ a_N)}$$
$$\times \prod_{j=1}^{N-1} p_j^{\delta(a_r,\ a_j)},$$

where $\delta(a_i,\ a_j) = 1$ for $i = j$ and $\delta(a_i,\ a_j) = 0$ for $i \neq j$.

2.46*. The model $\mathscr{F} = \{F(x;\ \theta),\ \theta \in \Theta\}$ is said to be *exponential* if the function $f(x;\ \theta)$ has the form

$$f(x;\ \theta) = \exp\ \{A(\theta)B(x) + C(\theta) + D(x)\}.$$

Prove that an efficient estimator τ^* for a parametric function $\tau(\theta)$ exists if and only if the model $\mathscr{F}$ is exponential and

$$\tau(\theta) = -\frac{C'(\theta)}{A'(\theta)},\quad \tau^* = \frac{1}{n}\sum_{i=1}^{n} B(X_i)$$

if θ is a scalar, and

$$\tau(\theta) = -\sum_{j=1}^{r} \frac{\partial C(\theta)}{\partial \theta_j} \Big/ \frac{\partial A(\theta)}{\partial \theta_j},\quad \tau^* = \frac{r}{n}\sum_{i=1}^{n} B(X_i)$$

if $\boldsymbol{\theta} = (\theta_1,\ \ldots,\ \theta_r)$.

Derive the formulas

$$\mathbf{D}_\theta \tau^* = \frac{\tau'(\theta)}{nA'(\theta)}$$

if θ is a scalar, and

$$\mathbf{D}_\theta \tau^* = \frac{1}{n}\sum_{j=1}^{n} \frac{\partial \tau(\theta)}{\partial \theta_j} \Big/ \frac{\partial A(\theta)}{\partial \theta_j}$$

if $\boldsymbol{\theta} = (\theta_1,\ \ldots,\ \theta_r)$.

Hint. Use the efficiency test.

2.47. Prove that for an exponential model with a scalar parameter the information function is

$$i(\theta) = (C'(\theta)A''(\theta) - C''(\theta)A'(\theta))/A'(\theta)$$

and

$$\mathbf{E}_\theta B(\xi) = -C'(\theta)/A'(\theta).$$

Hint. Compare the expression for $\mathbf{D}_\theta \tau^*$ in the previous problem with the Cramér-Rao bound.

2.48. Show that in the listed regular models the function $\tau(\theta)$ has an efficient estimator τ^* and variance $\mathbf{D}_\theta \tau^*$ as given in the following table:

Model	$\tau(\theta)$	τ^*	$\mathbf{D}_\theta \tau^*$
$\mathcal{N}(\theta, \sigma^2)$	θ	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	σ^2/n
$\mathcal{N}(\mu, \theta^2)$	θ^2	$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$	$2\theta^4/n$
$\Gamma(\theta, \lambda)$	θ	$\overline{X}/\lambda$	$\theta^2/\lambda n$
$\Gamma(a, \theta)$	$\Gamma'(\theta)/\Gamma(\theta)$	$\frac{1}{n}\sum_{i=1}^{n} \ln X_i - \ln a$	$\tau'(\theta)/n$
$B(\theta, 1)$	$1/\theta$	$-\frac{1}{n}\sum_{i=1}^{n} \ln X_i$	$1/(n\theta^2)$
$Bi(k, \theta)$	θ	$\overline{X}/k$	$\theta(1 - \theta)/kn$
$\Pi(\theta)$	θ	$\overline{X}$	θ/n
$\overline{Bi}(r, \theta)$	$r\theta/(1 - \theta)$	$\overline{X}$	$r\theta/[n(1 - \theta)^2]$

Hint. Use Problem 2.46.

2.49. Show that the sample mean $\overline{X}$ in a logistic model (see Problem 2.27) is not an efficient estimator for θ.

Hint. Use the result of Problem 2.27.

2.50. Prove that the estimator T^* in Problem 2.13 is optimal.

Hint. Consider the linear combinations of the form

$$\frac{1}{L}\left[a(\theta)\frac{\partial L}{\partial \theta}+b(\theta)\frac{\partial^2 L}{\partial \theta^2}\right]$$

and use the Bhattacharyya test.

2.51. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$, estimate the function $\tau(\theta) = \theta^2$ in the model $\Gamma(\theta, \lambda)$. Prove that $T^* = T^*(\mathbf{X}) = \dfrac{n}{\lambda(\lambda n + 1)}\overline{X}^2$ is an optimum unbiased estimator for $\tau(\theta)$. Compute $\mathbf{D}_\theta T^*$ and make sure that this estimator is not efficient.

Hint. Use Problems 2.21, 2.43, and the hint to Problem 2.50.

2.52. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathcal{N}(\theta_1, \theta_2^2)$. Apply the Bhattacharyya test and prove that $\overline{X}$ and S'^2 (see Problem 2.1 (d)) are optimum unbiased estimators for θ_1 and θ_2^2, respectively. Compare the variances of these estimators with the respective Cramér-Rao bounds.

Hint. In the first case it is sufficient to consider $\dfrac{\partial \ln L}{\partial \theta_1}$, in the second the linear combinations of the form $\dfrac{1}{L}\left[a(\theta)\dfrac{\partial L}{\partial \theta_2}+ b(\theta)\dfrac{\partial^2 L}{\partial \theta_1^2}\right]$. Use Problem 2.44.

2.53*. Let $\overline{X_1}$, $S_1'^2$ and $\overline{X_2}$, $S_2'^2$ be optimum unbiased estimators for the mean and variance of the same normal distribution computed from two independent samples of sizes n_1 and n_2, respectively. What functions of these statistics are the best estimators for these parameters and take into account all the original information? Compare the new estimators with the original ones. Which are more exact?

Hint. Use Problems 2.52 and 2.14.

2.54*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from an *inverse Gauss distribution* defined by the density

$$f(x; \lambda, \mu) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2}\exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\},$$

$$x > 0, \quad \lambda > 0, \quad \mu \neq 0.$$

(1) Show that $\overline{X}$ is an optimum unbiased estimator for the parameter μ whether the parameter λ is known or not. Derive the relations $\mathbf{E}X_1 = \mu$, $\mathbf{D}X_1 = \mu^3/\lambda$.

(2) Find an efficient estimator for λ^{-1} when μ is known.

Hint. Use Problem 2.46 and the Bhattacharyya test.

2.55. Suppose that we are estimating a differentiable vector function $\tau(\theta) = (\tau_1(\theta), \ldots, \tau_m(\theta))$, $\theta = (\theta_1, \ldots, \theta_r)$. Show that for a regular model its arbitrary unbiased estimator $\mathbf{T} = (T_1(\mathbf{X}), \ldots, T_m(\mathbf{X}))$ obeys the *information inequality*

$$\mathbf{D}_\theta(\mathbf{T}) = \|\mathrm{cov}_\theta\,(T_i,\ T_j)\|_1^m \geqslant \mathbf{B}'(\theta)\mathbf{I}_n^{-1}(\theta)\mathbf{B}(\theta),$$

where $\mathbf{B}(\theta) = \left\|\dfrac{\partial\tau_i(\theta)}{\partial\theta_j}\right\|$, and the inequality $\mathbf{A}_1 \geqslant \mathbf{A}_2$ between the matrices having the same dimensionality implies that the matrix $\mathbf{A}_1 - \mathbf{A}_2$ is nonnegative definite. Specifically, for $\tau(\theta) \equiv \theta$ we have $\mathbf{D}_\theta(\mathbf{T}) \geqslant \mathbf{I}_n^{-1}(\theta)$.

Hint. Consider an arbitrary linear combination $c_1\tau_1(\theta) + \ldots + c_m\tau_m(\theta) = \mathbf{c}'\tau(\theta)$ whose unbiased estimator is $\mathbf{c}'\mathbf{T}$ and apply the Cramér-Rao inequality for scalar estimators.

2.56. Show that if an efficient estimator exists for a function $\tau(\theta)$, then it is a sufficient statistic. Thus, a sufficient statistic always exists for regular exponential models (see Problem 2.46) and has the form $T(\mathbf{X}) = \sum_{i=1}^{n} B(X_i)$ (which also follows from the factorization test).

2.57. Prove the completeness of the sufficient statistic $T_n = \sum_{i=1}^{n} X_i$ for a binomial model $Bi(k, \theta)$ (see Problem 2.48). Show that in this case unbiased estimators only exist for the polynomials $\tau(\theta) = \sum_{j=0}^{r} a_j\theta^j$ of degree $r \leqslant kn$, and then the optimum estimator is

$$\tau^* = \sum_{j=0}^{r} a_j(T_n)_j/(kn)_j.$$

Compare this result with Problems 2.5, 2.7, and 2.8.

Hint. Use the reproducibility of the distribution $Bi(k, \theta)$ (see Problem 1.39 (3)).

2.58. Prove the completeness of the sufficient statistic $T_n = \sum_{i=1}^{n} X_i$ for Poisson's model $\Pi(\theta)$ (see Problem 2.48). Show that the statistic $\tau^* = \sum_j a_j(T_n)_j/n^j$ is an optimum estimator for the power series $\tau(\theta) = \sum_{j\geqslant 0} a_j\theta^j$ which is convergent for all $\theta > 0$.

Hint. Use the reproducibility of $\Pi(\theta)$ (see Problem 1.39 (4) and Problem 2.9).

2.59. (Continued from Problem 2.58.) Construct optimum estimators for the functions

$$\tau(\theta) = e^{\theta(z-1)}, \quad \pi_k(\theta) = e^{-\theta}\theta^k/k!, \quad k = 0, 1, \ldots,$$

and

$$\tau_r(\theta) = \mathbf{P}_\theta(\xi \geqslant r), \quad r = 1, 2, \ldots.$$

2.60*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from a power series distribution defined by the probabilities

$$f(x; \theta) = a(x)\theta^x/f(\theta),$$

$$x = l, l+1, \ldots, f(\theta) = \sum_{x=l}^{\infty} a(x)\theta^x, \quad \theta \in \Theta,$$

where $\Theta = (0, R)$ and $R > 0$ is the convergence radius of the series $f(\theta)$.

(1) Show that in this model an efficient estimator only exists for the function $\tau(\theta) = \theta f'(\theta)/f(\theta)$ and has the form $\tau^* = \overline{X}$.

(2) Prove that $T_n = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic with the distribution

$$\mathbf{P}_\theta(T_n = t) = \theta^t b_n(t)/f^n(\theta), \quad t \geqslant nl,$$

where $b_n(t) = \operatorname{coef}_{z^t} f^n(z)$.

(3) Make sure that the statistic

$$\tau_s^* = b_n(T_n - s)b_n^{-1}(T_n)I(T_n \geqslant nl + s)$$

is an optimum estimator for the function $\tau_s(\theta) = \theta^s$ for any $s = 1, 2, \ldots$, and

$$\tau^*(s) = b_{n-1}(T_n - s)b_n^{-1}(T_n)I(T_n \geqslant (n-1)l + s)$$

is an optimum estimator for $\tau(s; \theta) = \theta^s/f(\theta)$.

(4) Construct an optimum estimator for the function $\tau(\theta) = \sum_{j=r}^{\infty} a_j\theta^j$, where the power series is convergent on Θ. Show that $f^* = b_{n+1}(T_n)b_n^{-1}(T_n)I(T_n \geqslant (n+1)l)$ is an optimum estimator for the function $f(\theta)$.

Hints. (1) Apply the efficiency test for an exponential model (see Problem 2.46).

(2) Use the generating function

$$\varphi(z;\,\theta) \equiv \sum_{x=l}^{\infty} z^x f(x;\,\theta) = f(z\theta)/f(\theta).$$

(3) Take into account the relation $\sum_{j=l}^{k-nl} a(j)b_n(k-j) = b_{n+1}(k)$ for $k \geqslant (n+1)l$.

2.61. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from Poisson's distribution truncated at zero (see Problem 2.10), show that an optimum estimator for $\tau(\theta) = \theta$ is the statistic $\tau^* = T\Delta^n 0^{T-1}/\Delta^n 0^T$ for $T = X_1 + \ldots + X_n \geqslant n+1$, and the statistic $\tau^* = 0$ for $T = n$, where $\Delta^n 0^k = \sum_{r=0}^{n} (-1)^{n-r} C_n^r r^k$.

Hint. Use Problem 2.60.

2.62. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $\overline{Bi}(r, \theta)$, construct optimum estimators for $\tau_1(\theta) = \theta^s$ for integer $s \geqslant 1$, $\tau_2(\theta) = \mathbf{P}_\theta(\xi = 0) = (1-\theta)^r$, and $\tau_3(\theta) = \theta^s(1-\theta)^j$, $s \geqslant 0$, $j < rn$.

Hint. Use the solution to Problem 2.60 and the hint to Problem 2.11.

2.63*. Consider a model with a finite number N of possible outcomes and unknown probabilities $p_1, \ldots, p_N$ of the outcomes (see Problem 2.29). Show that $\mathbf{T} = (\nu_1, \ldots, \nu_{N-1})$ is a minimal complete sufficient statistic. Prove that unbiased estimators only exist in this model for the polynomials in $p_1, \ldots, p_N$ of a degree smaller than or equal to n and find the explicit form of these estimators.

Hint. Use the test for an r-parametric exponential family and Problems 2.29, 2.45, and 1.52 (b).

2.64. Prove that the estimators in Problems 2.13 and 2.16 are optimal.

Hint. Use the property of complete sufficient statistics.

2.65. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathscr{L}(\xi) = \mathscr{N}(\theta, \sigma^2)$.

(1) Construct an optimum estimator for $\tau(\theta) = \mathbf{P}_\theta(\xi \leqslant x_0)$, where x_0 is a given number.

(2) Show that an optimum estimator for the theoretical density $f_{\sigma^2}(x;\,\theta)$ has the form $f^*(x) = f_{\sigma_1^2}(x;\,\overline{X})$, where $\sigma_1^2 = (n-1)\sigma^2/n$. Specifically, it follows that if $\tau(\theta;\,\sigma^2) = \mathbf{E}_\theta\varphi(\xi)$, then its optimum estimator is $\tau^* = \int_{-\infty}^{\infty} \varphi(x)f^*(x)\,dx = \tau(\overline{X};\,\sigma_1^2)$. Apply this result to the functions $\varphi_1(x) = e^{itx}$, $\varphi_2(x) = x^2$, $\varphi_3(x) = I(x \leqslant x_0)$.

Hints. (1) Consider the unbiased estimator $T_1 = I(X_1 \leqslant x_0)$ and use the Rao-Blackwell-Kolmogorov theorem (see the solution to Problems 2.64 and 1.56).

(2) Check whether $f^*(x)$ satisfies the unbiasedness equation.

2.66. Using the test for an r-parametric exponential family, verify that the pairs $\left(\overline{X}, \sum_{i=1}^{n} X_i^2\right)$ and $(\overline{X}, S^2)$ are minimal complete sufficient statistic for the model $\mathcal{N}(\theta_1, \theta_2^2)$. Prove the optimality of the estimators in Problem 2.20 (compare with Problem 2.52).

Hint. Consider the new parameters $\theta_1' = \theta_1/\theta_2^2$, $\theta_2' = -1/(2\theta_2^2)$.

2.67. Show that the pair $\mathbf{T} = (\overline{X}, S^2)$ is a sufficient statistic for the model $\mathcal{N}(\theta, \gamma^2\theta^2)$ but this statistic is incomplete.

Hint. Consider the function $\varphi(\mathbf{T}) = (n + \gamma^2)S^2/[(n-1)\gamma^2]^{-1} - \overline{X^2}$ and calculate its mean.

2.68. Given $n \geqslant 2$ independent measurements of the diameter θ_1 of a circle, construct an optimum unbiased estimate for its area.

Hint. Assume that the measurement errors are $\mathcal{N}(0, \theta_2^2)$-normal random variables and use Problem 2.66.

2.69*. Prove the following assertion (*Basu theorem*): *if a complete sufficient statistic T exists for the model $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ and the statistic T_1 has a distribution independent of the parameter θ, then T_1 and T are independent.*

Hint. Show that for any event A the conditional $\mathbf{P}_\theta(T_1 \in A \mid T)$ and unconditional $\mathbf{P}_\theta(T_1 \in A)$ probabilities coincide.

2.70*. Let (X_1, X_2, X_3) be a sample from the distribution $\mathcal{L}(\xi) = \mathcal{N}(0, \theta^2)$. Construct an optimum estimator for $\tau(\theta) = \mathbf{P}_\theta(\xi \leqslant x_0)$.

Hint. See the hint to Problems 2.65, 2.69, and the solution to Problem 1.58.

2.71*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathcal{N}(\theta_1, \theta_2^2)$. Prove that the statistics $\mathbf{T} = (\overline{X}, S^2)$ and $\mathbf{U} = \left(\frac{X_i - \overline{X}}{S}, i = 1, \ldots, n\right)$ are independent.

Hint. Show that the distribution of $\mathbf{U}$ is independent of $\theta = (\theta_1, \theta_2)$ and apply the Basu theorem (see Problem 2.69).

2.72*. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $\mathcal{N}(\theta_1, \theta_2^2)$, construct an optimum unbiased estimator for the function

$$\tau(\theta) = \mathbf{P}_\theta(\xi \leqslant x_0) = \Phi\left(\frac{x_0 - \theta_1}{\theta_2}\right).$$

Hint. Consider an unbiased estimator $T_1 = I(X_1 \leqslant x_0)$ and calculate $H(\mathbf{T}) = \mathbf{E}_\theta(T_1|\mathbf{T})$, where $\mathbf{T} = (\overline{X}, S^2)$. Use Problems 2.71 and 1.58.

2.73. Make sure that the estimators in Problem 2.21 are optimal. Show that unbiased estimators for $\tau_a(\theta) = \theta^{-a}$ do not exist when $a - \lambda n \geqslant 0$ is an integer.

Hint. Use the completeness of a sufficient statistic T.

2.74*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\Gamma(\theta, \lambda)$, $T = X_1 + \ldots + X_n$, and $\varphi(x)$ be a given function for which $\tau(\theta) = \mathbf{E}_\theta \varphi(\xi)$ exists. Prove that an optimum estimator for $\tau(\theta)$ has the form

$$\tau^* = \frac{\Gamma(\lambda n)}{\Gamma(\lambda)\Gamma(\lambda(n-1))} \int_0^1 \varphi(xT) x^{\lambda - 1}(1 - x)^{(n-1)\lambda - 1}\, dx.$$

Derive the results of Problems 2.48, 2.51, and 2.73.

Hint. Prove that $\mathbf{E}_\theta \tau^* = \tau(\theta)$.

2.75*. (Continued from Problem 2.74.) Check whether the statistic

$$\tau^* = [1 - \mathrm{B}(t/T;\ \lambda,\ \lambda(n-1))] I(T \geqslant t)$$

is an optimum estimator for the *reliability function* $\tau(\theta;\ t) = \mathbf{P}_\theta(\xi \geqslant t)$, where $\mathrm{B}(x;\ a,\ b)$ is the function of a beta distribution $B(a,\ b)$. Specifically, for the distribution $\Gamma(\theta,\ 1)$ $\tau(\theta;\ t) = e^{-t/\theta}$ and $\tau^* = (1 - t/T)^{n-1} I(T \geqslant t)$.

Hint. Use Problem 2.74 assuming that $\varphi(x) = I(x \geqslant t)$.

2.76*. Prove that $T = T(\mathbf{X}) = \sum_{i=1}^{n} X_i^\lambda$ is a complete sufficient statistic for Weibull's distribution $W(0,\ \lambda,\ \theta)$ with an unknown scale parameter θ, and the optimum estimator for $\tau(\theta) = \mathbf{E}_\theta \varphi(\xi)$, where $\varphi(x)$ is a given function, has the form

$$\tau^* = (n - 1) \int_0^1 \varphi((tT)^{1/\lambda})(1 - t)^{n-2}\, dt.$$

Specifically, we have $\mathbf{E}_\theta \xi^\lambda = \theta^\lambda$, and therefore T/n is an optimum estimator for θ^λ.

2.77. Show that the pair $\mathbf{T} = (X_{(1)}, \overline{X})$ is a sufficient statistic for the *bivariate exponential distribution* $W(\theta_1,\ 1,\ \theta_2)$. Construct unbiased estimators of the form $\alpha X_{(1)} + \beta \overline{X}$ for the unknown parameters of the model and compute their variances.

Remark. Since $\mathbf{T}$ is a complete statistic, the respective estimators are optimal.

Hint. Apply the factorization test and use the solution to Problem 1.34 taking into account that $\mathcal{L}_\theta\left(\frac{\xi - \theta_1}{\theta_2}\right) = \Gamma(1, 1)$.

2.78. Let the observable random variable ξ have a range $[a(\theta), b]$, where $a(\theta)$ is a given monotone function of θ. Show that the minimal value of the sample $X_{(1)}$ is a sufficient statistic for θ if and only if the distribution density $f_\xi(x; \theta)$ has the form $f_\xi(x; \theta) = g(x)/h(\theta)$, $a(\theta) \leqslant x \leqslant b$. This result is also true for the statistic $X_{(n)}$ if the range is $[a, b(\theta)]$, where $b(\theta)$ is a given monotone function of θ.

Hint. Apply the factorization test.

2.79. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $R(0, \theta)$. Prove that $X_{(n)} = \max\limits_{1 \leqslant i \leqslant n} X_i$ is a complete sufficient statistic for θ. Prove that $T^* = \frac{n+1}{n} X_{(n)}$ is an optimum estimator for θ and, generally, $\tau^* = \tau(X_{(n)}) + \frac{1}{n} X_{(n)}\tau'(X_{(n)})$ is an optimum estimator for an arbitrary differentiable function $\tau(\theta)$. Consider the class of statistics $T_\lambda = \lambda T^*$ and show that it contains estimates with smaller standard error than that of T^*.

Hint. Use Problem 2.24.

2.80. Prove the completeness of a sufficient statistic $\mathbf{T} = (X_{(1)}, X_{(n)})$ for the model $R(\theta_1, \theta_2)$. Make sure that the estimators in Problem 2.25 are optimal. Construct optimum estimators for θ_1 and θ_2.

Hint. Use Problem 1.36.

2.81*. Show that $\mathbf{T} = (X_{(1)}, X_{(n)})$ is a sufficient statistic for the model $R(a(\theta), b(\theta))$, where $a(\theta) < b(\theta)\ \forall\theta$ are given continuous functions of the scalar parameter θ. Find the conditions for a univariate sufficient statistic to exist and establish its form. Verify that $\max(|X_{(1)}|, |X_{(n)}|)$ is a sufficient statistic for the model $R(-\theta, \theta)$, and $\mathbf{T}$ is a minimal sufficient statistic for the models $R(\theta, \theta + 1)$ and $R(\theta, 2\theta)$.

2.82*. Suppose that we have made one observation X on a discrete random variable distributed as

$$f(x; \theta) = \begin{cases} \theta & \text{for } x = -1, \\ \theta^x(1-\theta)^2 & \text{for } x = 0, 1, 2, \ldots, \end{cases} \qquad \theta \in (0, 1).$$

Show that X is a boundedly complete sufficient statistic.

Hint. Solve the unbiasedness equation $\mathbf{E}_\theta\varphi(X) = 0\ \forall\theta$ in the class of all functions and the subclass of bounded functions.

2.83*. *Estimation of the size of a finite population*. Let $m = 1$ in Problems 2.36 and 2.37. Prove that the random variable η is a complete sufficient statistic and, consequently, the estimators in Problem 2.37 are optimal.

Remark. This result is true for arbitrary m.

Maximum Likelihood Estimates

2.84*. Show that if an efficient estimator τ^* exists for a differentiable parametric function $\tau(\theta)$ in a regular model, then the m.l.e. $\hat{\theta}_n$ for the parameter θ is uniquely defined by the equation $\tau(\theta) = \tau^*$. Apply this result to find $\hat{\theta}_n$ for the models of Problem 2.48.

Hint. Use the efficiency test and show that $\left.\dfrac{\partial^2 \ln L}{\partial\theta^2}\right|_{\theta = \hat{\theta}_n} < 0$ (assume that the likelihood function $L = L(\mathbf{x}; \theta)$ is twice differentiable with respect to θ).

2.85. Compute the asymptotic efficiency of the sample median $T_n = X_{([n/2]+1)}$ which is an estimator for the mean θ in the model $\mathcal{N}(\theta, \sigma^2)$.

Hint. Use Problem 1.32 on the asymptotic normality of sample quantiles.

2.86. Prove that in the general normal model $\mathcal{N}(\theta_1, \theta_2^2)$ the m.l.e. $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n}) = (\overline{X}, S)$.

Hint. Write the likelihood equations and solve them.

2.87. (Continued from Problem 2.86.) Show that $\hat{\tau}_n = \Phi\left(\dfrac{x_0 - \overline{X}}{S}\right)$ is a m.l.e. for the function $\tau(\theta) = \Phi\left(\dfrac{x_0 - \theta_1}{\theta_2}\right)$ (see Problem 2.72).

Find an asymptotic distribution for $\hat{\tau}_n$ as $n \to \infty$.

Hint. Use the invariance of m.l.e.'s and the assertion that they are asymptotically normal.

2.88. Prove that the m.l.e. $\hat{\theta}_n$ for the parameter θ in the model $\mathcal{N}(\mu, \theta^2)$ is asymptotically unbiased and consistent (compare with Problem 2.16). Investigate its limiting distribution as $n \to \infty$. Calculate the asymptotic efficiency of the m.l.e. in Problem 2.15.

Hint. Use Problems 2.43 and 2.84.

2.89. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathcal{N}(\theta, 2\theta)$. Find the m.l.e. $\hat{\theta}_n$ and prove its consistency.

2.90. Given a sample $((X_1, Y_1), \ldots, (X_n, Y_n))$ from the bivariate

normal distribution $\mathcal{N}\left((0,\ 0),\ \begin{Vmatrix}\sigma^2 & \varrho\sigma^2\\ \varrho\sigma^2 & \sigma^2\end{Vmatrix}\right)$ with unknown $\sigma^2 > 0$ and $\varrho \in (-1,\ 1)$, construct the m.l.e.'s $\hat{\sigma}^2$ and $\hat{\varrho}$.

Hint. Go over to new parameters $\mathbf{q} = (q_1,\ q_2)$ putting $q_1 = q_1(\theta) = -\dfrac{1}{2\sigma^2(1-\varrho^2)}$, $q_2 = q_2(\theta) = \dfrac{\varrho}{\sigma^2(1-\varrho^2)}$ (here $\theta = (\sigma^2,\ \varrho)$) and use the invariance of the m.l.e.

2.91*. Given a sample $((X_1,\ Y_1),\ \ldots,\ (X_n,\ Y_n))$ from the bivariate normal distribution $\mathcal{N}\left((0,\ 0),\ \begin{Vmatrix}1 & \theta\\ \theta & 1\end{Vmatrix}\right)$, $\theta \in (-1,\ 1)$, write the likelihood equation to find $\hat{\theta}_n$ and calculate its asymptotic variance.

2.92. (Continued from Problem 2.91.) Assume that the sample correlation coefficient $T_n = \dfrac{1}{n} \sum_{i=1}^{n} X_i Y_i$ is an estimator for θ and compute its asymptotic efficiency.

Hint. Use the characteristic function to find the moments.

2.93*. Let $\mathbf{X} = (X_1,\ \ldots,\ X_n)$ be a sample from the k-variate normal distribution $\mathcal{N}(\mu,\ \Sigma)$ with unknown $\mu = (\mu_1,\ \ldots,\ \mu_k)$ and $\Sigma = \|\sigma_{ij}\|_1^k$, $|\Sigma| \neq 0$, i.e., $\mathbf{X}_l = (X_{l1},\ \ldots,\ X_{lk})$, $l = 1,\ \ldots,\ n$, are independent variables with the density

$$f(\mathbf{x};\ \theta) = \frac{1}{[(2\pi)^k|\Sigma|]^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\},$$

$\mathbf{x} = (x_1,\ \ldots,\ x_k)$, $\theta = (\mu,\ \Sigma)$. We write $\overline{\mathbf{X}} = (\overline{X}_1,\ \ldots,\ \overline{X}_k)$, where $\overline{X}_i = \dfrac{1}{n} \sum_{l=1}^{n} X_{li}$, $\hat{\Sigma} = \|S_{ij}\|_1^k$, with the sample covariance $S_{ij} = \dfrac{1}{n} \sum_{l=1}^{n} (X_{li} - \overline{X}_i)(X_{lj} - \overline{X}_j)$ which corresponds to the theoretical covariance σ_{ij} so that

$$\overline{\mathbf{X}} = \frac{1}{n}\sum_{l=1}^{n}\mathbf{X}_l,\quad \hat{\Sigma} = \hat{\Sigma}(\mathbf{X}) = \frac{1}{n}\sum_{l=1}^{n}(\mathbf{X}_l - \overline{\mathbf{X}})(\mathbf{X}_l - \overline{\mathbf{X}})'.$$

(1) Prove that $\mathbf{X}$ and $\hat{\Sigma}$ are m.l.e.'s for the parameters μ and Σ, respectively.

(2) Make sure that $\dfrac{n}{n-1}\hat{\Sigma}$ is an unbiased estimate for Σ.

(3) Obtain the expression

$$\max_{\theta} L(\mathbf{x};\ \theta) = L(\mathbf{x};\ \overline{\mathbf{x}},\ \hat{\Sigma}(\mathbf{x})) = (2\pi e)^{-kn/2}|\hat{\Sigma}(\mathbf{x})|^{-n/2}$$

for the maximum of the likelihood function.

| *Hint.* Reduce the likelihood function to the form

$$L(\mathbf{x};\ \theta) = [(2\pi)^k |\Sigma|]^{-n/2} \times \exp\left\{-\frac{n}{2}(\overline{\mathbf{x}} - \mu)'\Sigma^{-1}(\overline{\mathbf{x}} - \mu) - \frac{n}{2}\operatorname{tr}(\Sigma^{-1}\hat{\Sigma}(\mathbf{x}))\right\}$$

and use Problem 2.4.

2.94*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from a *lognormal distribution*, i.e., $X_i = e^{Y_i}$, where $\mathscr{L}(Y_i) = \mathscr{N}(\theta_1, \theta_2^2)$. Construct the m.l.e.'s for the functions $\tau_1(\theta) = \mathbf{E}_\theta X_1$ and $\tau_2(\theta) = \mathbf{D}_\theta X_1$. Compute $\mathbf{E}_\theta \hat{\tau}_{1n}$ and show that the estimate $\hat{\tau}_{1n}$ is asymptotically unbiased.

| *Hint.* Use the invariance of the m.l.e.

2.95. *Kapteyn's distribution.* This distribution is defined by the density

$$f(x;\ \theta) = \frac{g'(x)}{\sqrt{2\pi}\theta_2} \exp\left\{-\frac{1}{2\theta_2^2}(g(x) - \theta_1)^2\right\}, \quad \theta = (\theta_1, \theta_2),$$

where $g(x)$ is a differentiable monotone increasing function. Show that the following generalization of the result of Problem 2.86 is true, i.e., the m.l.e. $\hat{\theta}_n = (\overline{g}, T)$, where $\overline{g} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$, $T^2 = \frac{1}{n}\sum_{i=1}^{n}(g(X_i) - \overline{g})^2$.

Is $\overline{g}$ an efficient estimate for θ_1? Show that the statistic

$$T_1^2 = \frac{1}{n}\sum_{i=1}^{n}(g(X_i) - a)^2$$

is an efficient estimate for θ_2^2 when $\theta_1 = a$, where a is known (compare with the respective results for the normal model in Problem 2.48).

| *Hint.* Use Problem 2.46.

2.96. Suppose that a random variable ξ has a power series distribution (see Problem 2.60). Show that here the likelihood equation for finding the m.l.e. $\hat{\theta}_n$ has the form $\mu(\theta) = \overline{X}$, where $\mu(\theta) = \mathbf{E}_\theta \xi$. Calculate the asymptotic variance of $\hat{\theta}_n$. Apply the results to estimate the parameter θ in the model $\overline{Bi}(r, \theta)$.

2.97. Write the accumulation method equations to compute approximately the m.l.e. $\hat{\theta}_n$ for the parameter θ of Poisson's distribution truncated at zero (see Problem 2.10).

| *Hint.* Use the solution to Problem 2.96.

2.98. Suppose that in a polynomial distribution $M(n;\ p_1, \ldots, p_N)$ the probabilities of the outcomes are $p_i = p_i(\theta)$, $i = 1, \ldots, N$, where θ is an unknown scalar parameter. Write the accumulation method equations for an approximate calculation of the m.l.e. $\hat{\theta}_n$.

2.99. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$, estimate the parameter θ of the Cauchy model $C(\theta)$. Using the accumulation method, write the equations to calculate the m.l.e. $\hat{\theta}_n$ approximately. Use the sample median $T_n = X_{\left(\left[\frac{n}{2}\right]+1\right)}$ as an estimator for θ and compute its asymptotic efficiency.

| *Hint.* Use Problems 2.43 and 1.32.

2.100. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the uniform distribution $R(0, \theta)$. Show that here the m.l.e. $\hat{\theta}_n = X_{(n)}$, make sure that it is consistent, and find its limiting distribution as $n \to \infty$.

| *Hint.* Use Problems 2.24 and 2.79.

2.101. Show that in the model $R(\theta - 1/2, \theta + 1/2)$ any $\theta \in [X_{(n)} - 1/2, X_{(1)} + 1/2]$ is a m.l.e. $\hat{\theta}_n$. What point of this interval is an unbiased estimate for θ?

| *Hint.* Use the solutions to Problems 2.80 and 1.36.

2.102. Show that for the shift parameter θ in Weibull's distribution $W(\theta, \alpha, b)$, $0 < \alpha \leqslant 1$, the m.l.e. $\hat{\theta}_n$ is $X_{(1)}$, prove that it is consistent, and find its limiting distribution as $n \to \infty$.

| *Hint.* Use the solutions to Problems 1.37 and 2.26.

2.103. A random variable ξ which describes the lifetimes of the elements of an electronic device has Rayleigh's distribution $W(0, 2, \sqrt{\theta})$ with the density $f(x; \theta) = (2x/\theta)e^{-x^2/\theta}$, $x \geqslant 0$. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$, construct the m.l.e. $\hat{\theta}_n$ (compare with Problem 2.76).

2.104. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $\Gamma(\theta, \lambda)$, estimate the function $\tau(\theta) = 1/\theta$. Show that the m.l.e. $\hat{\tau}_n = \lambda/\overline{X}$. Make sure that the estimate is consistent and find its limiting distribution as $n \to \infty$.

| *Hint.* Use Problems 2.21, 2.43, and 2.84.

2.105*. Prove that for the *Laplace distribution* defined by the density $f(x; \theta) = \dfrac{1}{2} e^{-|x-\theta|}$, $x \in R$, the m.l.e. $\hat{\theta}_n$ coincides with the sample median. Can we use here the theorem on the asymptotic normality of the m.l.e.?

2.106*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathcal{N}(\theta, 1)$. Then (see Problem 2.84) $\hat{\theta}_n = \overline{X}$ and $\mathcal{L}_\theta(\overline{X}) = \mathcal{N}(\theta, 1/n)$. Take the statistic

$$T_n = \overline{X} I(|\overline{X}| \geqslant a_n) + b\overline{X} I(|\overline{X}| < a_n)$$

as an estimator for θ, where $a_n \to 0$, but $\sqrt{n} a_n \to \infty$ as $n \to \infty$, and calculate its asymptotic efficiency.

2.107. Give examples of the m.l.e. $\hat{\theta}_n$ for which $\mathbf{D}_\theta \hat{\theta}_n = o(n^{-1})$.

| *Hint.* Consider the model $R(0, \theta)$ (see Problem 2.100) and Weibull's model (see Problems 2.102 and 1.37).

2.108. Estimate the function $\tau(\theta) = \theta^{-1}$ in the model $\Pi(\theta)$ and make sure that the m.l.e. $\hat{\tau}_n$ has no finite moments for any n but, nevertheless, its asymptotic variance exists and is equal to $(\theta^3 n)^{-1}$.

| *Hint.* Use Problems 2.84, 1.39, and 2.43.

2.109*. *Variance-stabilizing transformations.* For the models $Bi(k, \theta)$, $\Pi(\theta)$, $\mathcal{N}(\mu, \theta^2)$, and $\Gamma(\theta, \lambda)$ find the parametric functions $\tau(\theta)$ such that the asymptotic variances of the respective m.l.e.'s $\hat{\tau}_n$ are independent of the parameter θ.

| *Hint.* Use Problem 2.43.

2.110. Simulate samples of sizes $n = 10$, $n = 100$, $n = 1000$ and obtain the m.l.e.'s for the parameters of the following distributions:

(1) $\mathcal{N}(\theta_1, \theta_2^2)$ with $\theta_1 = 1$, $\theta_2^2 = 4$;

(2) $Bi(1, \theta)$ with $\theta = 0.7$;

(3) $R(0, \theta)$ with $\theta = 1$.

| *Hint.* Use Problems 2.86, 2.84, and 2.100, respectively.

2.111*. *Estimation of the size of a finite population.* Under the conditions of Problem 2.83 show that the m.l.e. $\hat{N}$ for an unknown parameter of the population N can be found in a unique way (for $\eta > 1$) from the condition

$$S(\hat{N}, \eta) \leqslant n < S(\hat{N} - 1, \eta),$$

where $S(N, k) = \ln \dfrac{N+1}{N+1-k} \Big/ \ln \dfrac{N+1}{N}$ for $N \geqslant k > 1$, $S(k-1, k) = \infty$. If $\eta = 1$, then $\hat{N} = 1$.

Find the values of η for which $\hat{N} = \eta$. Assume that $n, N \to \infty$, $0 < \alpha_0 \leqslant \alpha = \dfrac{n}{N+1} \leqslant \alpha_1 < \infty$, where α_0, α_1 are constants, and find an approximate expression for the m.l.e. $\hat{\alpha} = n/(\hat{N} + 1)$. Generalize the result to the case of an arbitrary m.

2.112. (Continued from Problem 2.111.) To estimate the unknown number N of fish in a lake, we carry out the following experiment. At the first stage we catch m_1 fish at random and without replacement, ring them, and release them to the lake. Using the same scheme at the second stage, we catch m_2 fish and register the number μ_2 of ringed fish (so that the total number of fish caught during the two stages is $\eta = m_1 + m_2 - \mu_2$). Show that given μ_2, the m.l.e. $\hat{N}$ is defined by $\hat{N} = \left[\dfrac{m_1 m_2}{\mu_2}\right]$, where $[\cdot]$ is an integer part. Compare this result for $m_1 = m_2$ with that obtained in Problem 2.36.

Hint. Take into account that the statistic μ_2 has a hypergeometric distribution $H(m_1, N, m_2)$.

2.113*. *Sampling inspection.* Suppose we have a batch of N products, which contains an (unknown) number D of defective items. In order to estimate D or a given function $\tau(D)$, we draw n $(n < N)$ items from the batch at random and without replacement. Each item is subjected to quality control. Let $X_i = 1$ if the ith tested item is defective, and $X_i = 0$ otherwise, $i = 1, \ldots, n$.

(1) Show that $d_n = X_1 + \ldots + X_n$ (the total number of defectives in the sample $\mathbf{X} = (X_1, \ldots, X_n)$) is a complete sufficient statistic for D and has a hypergeometric distribution $H(D, N, n)$. Having proved this, make sure that unbiased estimates exist only if $\tau(D)$ is a polynomial of a degree no greater than n. In this case if $\tau(D) = \sum_{j=0}^{n} a_j(D)_j$, $(D)_j = D(D-1)\ldots(D-j+1)$, $(D)_0 = 1$, then the statistic

$$\tau^* = T(d_n) = \sum_{j=0}^{n} a_j(d_n)_j(N)_j/(n)_j$$

is an optimum unbiased estimator for $\tau(d)$.

(2) Find the explicit form of optimum estimators for the functions $\tau_1(D) = D$ and $\tau_2(D) = D(N - D)$, which are, up to the multipliers, the mean and variance, respectively, of the statistic d_n (see Sec. 1.6).

(3) Make sure that the m.l.e. is $\hat{D}_n = [(N + 1)d_n/n]$.

Hint. Use the solution to Problem 2.33 and the formulas for the moments of $H(D, N, n)$.

2.114. *Grouping of statistical data.* Let $\mathbf{X}_j = (X_{j1}, \ldots, X_{jn}), j = 1, \ldots, k$, be independent samples from the respective distributions $\mathcal{N}(\theta_{j1}, \theta_2^2)$, $j = 1, \ldots, k$, and let $\overline{X}_j$, $S_j^2 = S^2(\mathbf{X}_j)$ be the respective sample means and variances. Prove that $\hat{\theta} = (\overline{X}_1, \ldots, \overline{X}_k, \hat{\theta}_2)$ is a m.l.e. for $\theta = (\theta_{11}, \ldots, \theta_{k1}, \theta_2)$, where $\hat{\theta}_2^2 = \dfrac{1}{n_1 + \ldots + n_k} \sum_{j=1}^{k} n_j S_j^2$, and that the statistic

$$\tilde{\theta}_2^2 = \frac{n_1 + \ldots + n_k}{n_1 + \ldots + n_k - k}\,\hat{\theta}_2^2 = \frac{1}{n_1 + \ldots + n_k - k} \sum_{j=1}^{k} n_j S_j^2$$

is an unbiased estimator for the common variance θ_2^2.

Hint. Use the solution to Problem 2.86.

Confidence Estimation

2.115. Show that given a sample $\mathbf{X} = (X_1, \ldots, X_n)$, the γ-confidence interval for the parameter θ in the model $\mathcal{N}(\theta, \theta^2)$, $\theta > 0$, has the form $(\overline{X}/(1 + c_\gamma/\sqrt{n}),\ \overline{X}/(1 - c_\gamma/\sqrt{n}))$*. Find a similar solution for the model $\mathcal{N}(\theta, \theta^2)$, $\theta < 0$.

| *Hint.* Use the fact that $\mathcal{L}_\theta((\overline{X} - \theta)\sqrt{n}/\theta) = \mathcal{N}(0, 1)$.

2.116. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathcal{N}(\theta, \sigma^2)$.

(1) Show that any interval of the form $\Delta_\gamma(\mathbf{X}) = \left(\overline{X} - \frac{\sigma}{\sqrt{n}} g_2,\ \overline{X} - \frac{\sigma}{\sqrt{n}} g_1\right)$, where $g_1 < g_2$ are any numbers satisfying the condition $\Phi(g_2) - \Phi(g_1) = \gamma$, is a γ-confidence interval for the parameter θ.

Prove that $\Delta_\gamma^*(\mathbf{X}) = \left(\overline{X} \pm \frac{\sigma}{\sqrt{n}} c_\gamma\right)$ is the shortest interval among the γ-confidence intervals.

(2) How many observations should be made ($n = n(l, \gamma)$) for the precision of the localization parameter to be equal to l for a given confidence level γ? Calculate $n(l, \gamma)$ for $\gamma = 0.99$, $l = 0.5$, and $l = 0.1$ ($\sigma = 1$). How does the confidence level γ change depending on l and n?

| *Hint.* Use the central statistic $G(\mathbf{X};\ \theta) = \frac{\sqrt{n}}{\sigma}(\overline{X} - \theta)$.

2.117. Prove that the γ-confidence interval for the mean-square deviation of θ in the model $\mathcal{N}(\mu, \theta^2)$ is any interval $\delta_\gamma(\mathbf{X}) = (T/a_2, T/a_1)$, where $T^2 = \sum_{i=1}^{n} (X_i - \mu)^2$, and the numbers $a_1 < a_2$ are chosen from the condition $\int_{a_1}^{a_2} x k_n(x^2)\, dx = \gamma/2$, where $k_n(t)$ is the density of the distribution $\chi^2(n)$. Define the shortest interval $\delta_\gamma^*(\mathbf{X})$ in this class.

| *Hint.* Use the fact that $\mathcal{L}_\theta(T^2/\theta^2) = \chi^2(n)$.

2.118. (Continued from Problem 2.117.) Show that the central γ-confidence interval for the variance of θ^2 has the form

$$\Delta_\gamma(\mathbf{X}) = (T^2/g_2,\ T^2/g_1), \qquad g_1 = \chi^2_{(1-\gamma)/2,n}, \quad g_2 = \chi^2_{(1+\gamma)/2,n},$$

---

* Recall that $c_\gamma = u_{(1+\gamma)/2} = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$.

while the interval $\Delta_\gamma^*(\mathbf{X}) = (T_1^{*2}, T_2^{*2})$ is the shortest among these intervals, where the numbers $g_1 < g_2$ satisfy the condition $\int_{g_1}^{g_2} k_n(t)\,dt = \gamma$ (see the solution to the previous problem).

2.119. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $\mathcal{N}(\theta_1, \theta_2^2)$, construct the one- and two-sided γ-confidence intervals for the mean θ_1.

Hint. Use the assertion

$$\mathcal{L}_\theta\left(\sqrt{n-1}\,\frac{\overline{X} - \theta_1}{S}\right) = S(n-1).$$

2.120. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $\mathcal{N}(\theta_1, \theta_2^2)$, construct the one- and two-sided γ-confidence intervals for the variance $\tau = \theta_2^2$.

Hint. Use Fisher's theorem.

2.121. Given the realization (2.96, 3.07, 3.02, 2.98, 3.06) of a sample of size $n = 5$ from a normal distribution with unknown parameters, construct 95% confidence intervals for the mean and variance.

2.122. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)$ be two independent samples, the first from the distribution $\mathcal{N}(\theta^{(1)}, \sigma_1^2)$, and the second from $\mathcal{N}(\theta^{(2)}, \sigma_2^2)$. Construct a γ-confidence interval for the difference $\tau = \theta^{(1)} - \theta^{(2)}$ of the means.

Hint. Show that $\mathcal{L}((\overline{X} - \overline{Y} - \tau)/\sigma) = \mathcal{N}(0, 1)$, $\sigma^2 = \dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}$.

2.123. (Continued from Problem 2.122.) In contrast to the previous case, all the observations have the same unknown variance θ_2^2, i.e., $\mathcal{L}(X_i) = \mathcal{N}(\theta_1^{(1)}, \theta_2^2)$, $\mathcal{L}(Y_j) = \mathcal{N}(\theta_1^{(2)}, \theta_2^2)$. Estimate the difference $\tau = \theta_1^{(1)} - \theta_1^{(2)}$ of the means. Consider a more general situation when the variances are unknown but only differ by a known factor, i.e., $\mathcal{L}(X_i) = \mathcal{N}(\theta_1^{(1)}, c\theta_2^2)$, $\mathcal{L}(Y_j) = \mathcal{N}(\theta_1^{(2)}, \theta_2^2)$, where c is known.

Hint. Show that the random variable

$$t_{m+n-2} = \sqrt{\frac{mn(m+n-2)}{m+n}}\,\frac{\overline{X} - \overline{Y} - \tau}{\sqrt{nS^2(\mathbf{X}) + mS^2(\mathbf{Y})}}$$

has Student's distribution $S(m + n - 2)$.

2.124. Two measurements at the same points of an angle gave (in degrees) 20.76 and 20.98. Six more such measurements of the same angle made by another device gave 21.64, 21.54, 22.32, 20.56, 21.43,

21.07. We assume that random measurement errors are normally distributed and that the first device is less accurate (the respective variance is four times that produced by the second device). Calculate a 95% confidence interval for the difference of the systematic errors due to the use of these devices.

| *Hint.* Use the solution to Problem 2.123.

2.125. (Continued from Problem 2.123.) Suppose that the samples have different variances, i.e., $\mathscr{L}(X_i) = \mathscr{N}(\theta_1^{(1)}, \theta_2^{(1)^2})$, $\mathscr{L}(Y_j) = \mathscr{N}(\theta_1^{(2)}, \theta_2^{(2)^2})$. Construct a confidence interval for the ratio $\tau = \theta_2^{(1)^2}/\theta_2^{(2)^2}$.

| *Hint.* Show that the central statistic here has the form

$$F_{n-1,m-1} = \frac{n(m-1)}{m(n-1)} \frac{S^2(\mathbf{X})}{S^2(\mathbf{Y})} \Big/ \tau.$$

2.126. Two laboratories measured the concentration (in %) of sulphur in a standard sample of diesel fuel. Six independent measurements in the first laboratory gave 0.869, 0.874, 0.867, 0.875, 0.870, 0.869. Five similar measurements in the second laboratory gave 0.865, 0.870, 0.866, 0.871, 0.868. Under the assumption that the measurement errors are normally distributed, construct a 90% confidence interval for the ratio of the measurements of variances in the two laboratories. If there are grounds to assume that the variances are the same, construct a similar interval for the difference of systematic errors in both laboratories.

2.127. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)$ be samples from the distributions $\Gamma(\theta_1, 1)$ and $\Gamma(\theta_2, 1)$, respectively. Construct a central γ-confidence interval for the ratio $\tau = \theta_2/\theta_1$.

| *Hint.* Use Problem 1.51.

2.128. Make sure that $\left(X_{(1)} + \dfrac{\ln(1-\gamma)}{n}, X_{(1)}\right)$, where $X_{(1)} = \min\limits_{1 \leqslant i \leqslant n} X_i$, is a γ-confidence interval for the parameter θ of an exponential distribution with the density $f(x; \theta) = e^{-(x-\theta)}$, $x \geqslant \theta$. What is the central γ-confidence interval?

| *Hint.* Find the distribution of the statistic $X_{(1)}$ and take into account that $\{X_{(1)} \geqslant \theta\}$ is a true event.

2.129. Given a sample of size n, make sure that $(X_{(n)}, X_{(n)}/\sqrt[n]{1-\gamma})$ is a γ-confidence interval for the parameter θ in the model $R(0, \theta)$ (as in Problem 2.128).

| *Hint.* Show that $\mathscr{L}_\theta((X_{(n)}/\theta)^n) = R(0, 1)$ (see Problem 1.35).

2.130. Consider the model $W(0, \lambda, \theta)$ (see Problem 2.76). Show that the interval $(2T/\chi^2_{(1+\gamma)/2,2n}, 2T/\chi^2_{(1-\gamma)/2,2n})$ is a central γ-confidence interval for the function $\tau(\theta) = \theta^\lambda$. Specifically, for $\lambda = 1$ we have the solution for an exponential model $\Gamma(\theta, 1)$.

| *Hint.* Use the solution to Problem 2.76.

2.131*. Show that the γ-confidence region for the parameters $(\theta_1, \tau = \theta_2^2)$ in the general normal model $\mathcal{N}(\theta_1, \theta_2^2)$ found from the sample $\mathbf{X} = (X_1, \ldots, X_n)$ has the form

$$\mathcal{G}_\gamma(\mathbf{X}) = \{(\theta_1; \tau): \tau > n(\overline{X} - \theta_1)^2/c^2_{\gamma_1},$$
$$nS^2/\chi^2_{(1+\gamma_2)/2,n-1} < \tau < nS^2/\chi^2_{(1-\gamma_2)/2,n-1}\},$$

where $\gamma_1\gamma_2 = \gamma$.

| *Hint.* Use Fisher's theorem.

2.132*. Let $(X_i = (X_{i1}, X_{i2}), i = 1, \ldots, n)$ be a sample from the bivariate normal distribution

$$\mathcal{N}\left((\theta_1, \theta_2), \quad \Sigma = \begin{Vmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho \\ \sigma_1\sigma_2\varrho & \sigma_2^2 \end{Vmatrix}\right), \quad -1 < \varrho < 1,$$

with a known matrix Σ. Using Problem 1.59, construct the γ-confidence region for $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

2.133. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from $Bi(1, \theta)$. Using a point estimator $T = \overline{X}$ for the parameter θ, show that the central γ-confidence interval (T_1, T_2) for it is defined by the conditions

$$\sum_{r=nT}^{n} C_n^r T_1^r(1 - T_1)^{n-r} = \sum_{r=0}^{nT} C_n^r T_2^r(1 - T_2)^{n-r} = \frac{1-\gamma}{2};$$

here $T_1 = Z\left(\frac{1-\gamma}{2}; nT, n - nT + 1\right)$, $T_2 = Z\left(\frac{1+\gamma}{2}; nT + 1, n - nT\right)$, where $Z(p; a, b)$ is a p-quantile of the beta distribution $B(a, b)$. Construct an approximate γ-confidence interval for θ for large n.

| *Hint.* Use Problems 1.39 (3), 2.43, and 2.84.

2.134. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the Bernoulli model $Bi(1, \theta)$, construct an asymptotic (for $n \to \infty$) γ-confidence interval for θ based on the normal approximation $\mathcal{L}_\theta(\sqrt{n}(\overline{X} - \theta)/\sqrt{\theta(1-\theta)}) \sim \mathcal{N}(0, 1)$ (the De Moivre-Laplace theorem). Compare the resultant solution with that based on the asymptotic properties of maximum likelihood estimates (see Problem 2.133).

2.135. (Continued from Problem 2.134.) Show that $\left(\arcsin\sqrt{\overline{X}} \pm \frac{c_\gamma}{2\sqrt{n}}\right)$ is an asymptotic γ-confidence interval for the function $\tau(\theta) = \arcsin\sqrt{\theta}$ and then find an approximate confidence interval for θ.

Hint. Use Problem 2.109.

2.136. In 540 Bernoulli trials a positive result was observed 216 times. Calculate a 95% confidence interval for the variance of the number of positive outcomes.

2.137. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\Pi(\theta)$. Using the point estimator $T = \overline{X}$, show that the central γ-confidence interval (T_1, T_2) for θ is defined by the conditions

$$\sum_{r=nT}^{\infty} e^{-nT_1}\frac{(nT_1)^r}{r!} = \sum_{r=0}^{nT} e^{-nT_2}\frac{(nT_2)^r}{r!} = \frac{1-\gamma}{2},$$

where $T_1 = \frac{1}{2n}\chi^2_{(1-\gamma)/2,\,2nT}$, $T_2 = \frac{1}{2n}\chi^2_{(1+\gamma)/2,\,2nT+2}$. An approximate γ-confidence interval for large n is $(\overline{X} \pm c_\gamma\sqrt{\overline{X}/n})$.

2.138. Construct an asymptotic γ-confidence interval for the parameter θ of Poisson's model $\Pi(\theta)$. Use the norml approximation $\mathcal{L}_\theta(2\sqrt{n}(\sqrt{\overline{X}} - \sqrt{\theta})) \sim \mathcal{N}(0, 1)$ (see Problem 2.109) or the approximation $\mathcal{L}_\theta\left(\frac{\sqrt{n}(\overline{X} - \theta)}{\sqrt{\theta}}\right) \sim \mathcal{N}(0, 1)$ (the Central Limit Theorem). Compare the result with that of the previous problem.

2.139. Independent random variables X_1 and X_2 have Poisson's distribution with the parameters λ_1 and λ_2, respectively. Suppose that we know their sum $X_1 + X_2 = n$. Construct a confidence interval for $\theta = \lambda_1/(\lambda_1 + \lambda_2)$ given an observation on X_1.

Hint. Find the conditional distribution $\mathcal{L}(X_1 | X_1 + X_2 = n)$ (see Problem 1.54) and use the solution to Problem 2.133.

2.140. Construct an asymptotic γ-confidence interval for the parameter θ of a power series distribution (see Problem 2.60). Use the result to estimate the parameter θ in the model $\overline{Bi}(r, \theta)$.

Hint. Use Problem 2.96 and its solution.

2.141. Construct an asymptotic γ-confidence interval for the parameter θ in the model $\Gamma(\theta, \lambda)$.

Hint. Use Problems 2.43 and 2.48, and the approximation in Problem 2.109.

2.142. Construct an asymptotic γ-confidence interval for the parameter θ in the model $\mathcal{N}(\mu, \theta^2)$.

Hint. Use the approximation $\mathcal{L}_\theta(\sqrt{2n}(\ln \hat{\theta}_n - \ln \theta)) \sim \mathcal{N}(0, 1)$ in Problem 2.109.

2.143. Construct an asymptotic γ-confidence interval for the function $\tau(\theta) = \Phi\left(\dfrac{x_0 - \theta_1}{\theta_2}\right)$ in the model $\mathcal{N}(\theta_1, \theta_2^2)$ (see Problem 2.72).

Hint. Use Problem 2.87.

2.144*. Given a polynomial model $M(n; p_1, \ldots, p_N)$ with the unknown parameters $p_1, \ldots, p_N$ (see Problem 2.29), construct an asymptotic (for $n \to \infty$) γ-confidence region for $p_1, \ldots, p_N$ based on the respective maximum likelihood estimates.

Hint. Use Problems 2.63, 2.45, and the asymptotic version of Problem 1.40, i.e., if $\mathcal{L}(\mathbf{Y}_n) \sim \mathcal{N}(\mu_n, \Sigma_n)$ as $n \to \infty$ and $|\Sigma_n| \neq 0$, then $\mathcal{L}((\mathbf{Y}_n - \mu_n)'\Sigma_n^{-1}(\mathbf{Y}_n - \mu_n)) \to \chi^2(m)$, where m is the dimensionality of the vector $\mathbf{Y}_n$.

2.145*. Let n, $\overline{X}$, and S^2 be the sample size, mean, and variance from the distribution $\mathcal{N}(\theta_1, \theta_2^2)$. Show that the result of the next, $(n + 1)$th, trial is the interval $(\overline{X} \pm t_{(1+\gamma)/2, n-1} S\sqrt{(n + 1)/(n - 1)})$ with probability γ.

Hint. Use Fisher's theorem.

2.146. (Continued from Problem 2.145.) Five independent measurements of a body gave (in grams) 4.12, 3.92, 4.55, 4.04, 4.35. Assume that the measurement errors are $\mathcal{N}(0, \theta_2^2)$-normally distributed random variables and construct a 95% confidence interval for the result of the next (sixth) measurement.

2.147. Given $\mathcal{L}(\xi) = \chi^2(n)$, where n is the unknown number of the degrees of freedom, construct an approximate 90% confidence interval for n, which corresponds to the realization $\xi = 157.4$.

Hint. Use the normal approximation for the χ^2-distribution (Problem 1.45).

2.148. Given the samples from Problem 2.110, construct γ-confidence intervals for the respective parameters.

Hint. Use Problems 2.119, 2.120, 2.133, and 2.129, respectively.

2.149. Let $X_1, \ldots, X_n$ be independent observations on a random variable ξ with $\mathbf{E}\xi^{2k} < \infty$. Prove that an asymptotic γ-confidence interval for the moment $\alpha_k = \mathbf{E}\xi^k$ has the form $(A_{nk} \pm c_\gamma\sqrt{(A_{n,2k} - A_{nk}^2)/n})$.

Hint. Using Problem 1.28, show that

$$\mathcal{L}(\sqrt{n}(A_{nk} - \alpha_k)/\sqrt{A_{n,2k} - A_{nk}^2}) \to \mathcal{N}(0, 1)$$

as $n \to \infty$.

2.150. Let ϱ_n be a sample correlation coefficient constructed from n observations on a two-dimensional random variable $\xi = (\xi_1, \xi_2)$ with unknown parameters and $Z_n = \frac{1}{2} \ln ((1 + \varrho_n)/(1 - \varrho_n))$. Using the normal approximation $\mathscr{L}(Z_n) \sim \mathscr{N}\left(\zeta, \frac{1}{n-3}\right)$, where $\zeta = \frac{1}{2} \ln \frac{1+\varrho}{1-\varrho}$, $\varrho = \operatorname{corr}(\xi_1, \xi_2)$, construct an asymptotic γ-confidence interval for ϱ.

Hint. Use the fact that the function $\tau(x) = \ln \frac{1+x}{1-x}$, $x \in (-1, 1)$, is monotone.

CHAPTER 3

Tests of Statistical Hypotheses

3.1. Any assumption on the form or properties of the random variables observed in an experiment is called a *statistical hypothesis* (or simply *hypothesis*). If a hypothesis H_0 (called the *null hypothesis*) is formulated for the process under investigation, then we test it by constructing a rule (algorithm) which allows us to accept or reject H_0 on the basis of observations (statistical data). Such rules are called the *goodness of fit tests* (or simply *tests*) for the hypothesis H_0. If H_0 corresponds to a unique distribution of observations, then it is called a *simple hypothesis*, orherwise it is a *composite hypothesis*.

Let the outcome of an experiment be defined by a random variable $\mathbf{X} = (X_1, \ldots, X_n)$, and let H_0 be a hypothesis about its distribution. We assume that the statistic $T = T(\mathbf{X})$ describes the deviation of the empirical data from the respective (under the hypothesis H_0) hypothetical values and its distribution is known (exactly or approximately) when H_0 is true. Then for any sufficiently small number $\alpha > 0$ we can define a subset $\mathscr{T}_{1\alpha} = \{t: t = T(\mathbf{x}), \mathbf{x} \in \mathscr{X}\}$, which satisfies (exactly or approximately) the condition

$$\mathbf{P}(T \in \mathscr{T}_{1\alpha} | H_0) \leqslant \alpha. \tag{3.1}$$

Any subset $\mathscr{T}_{1\alpha}$ brings about the following goodness of fit test for the hypothesis H_0: if $t = T(\mathbf{x})$ is the observed value of the statistic $T(\mathbf{x})$, then the hypothesis H_0 is rejected for $t \in \mathscr{T}_{1\alpha}$, otherwise we assume that the data are consistent with H_0, or, if $t \notin \mathscr{T}_{1\alpha}$, then the hypothesis H_0 is accepted (note that $t \notin \mathscr{T}_{1\alpha}$ does not prove that H_0 is true). If H_0 is true, we may, according to our rule, reject it (i.e., make a wrong decision) with the probability smaller than or equal to α. The number α is called the *significance level* of a test, and the set $\mathscr{T}_{1\alpha}$ is called the *critical set* (*region*) for the hypothesis H_0. The statistic T is called a *test statistic*, and the test itself is called the $\mathscr{T}_{1\alpha}$-*test*.

Thus, in this technique a test is defined by the critical region $\mathscr{T}_{1\alpha}$ in the range of the statistic T for a chosen significance level α. Differ-

ent tests (generated by different statistics T) can be compared using the notions of an alternative distribution (alternative hypothesis) and the power of a test.

Any admissible distribution $F_{\mathbf{X}} = F$ of a sample $\mathbf{X}$, which differs from the hypothetical (under the hypothesis H_0) distribution is called an *alternative distribution*, or *alternative*. The set of all the alternatives is called an *alternative hypothesis* and is denoted H_1. The *power function* of the $\mathscr{T}_{1\alpha}$-test is a functional

$$W(F) = W(\mathscr{T}_{1\alpha};\, F) \equiv \mathbf{P}(T \in \mathscr{T}_{1\alpha} | F) \tag{3.2}$$

on the set of all the admissible distributions $\{\mathscr{F}\}$. Thus, $W(F)$ is the probability that the values of the test statistic are in the critical region when F is the true distribution of the observations. If $F \in H_1$, the value of $W(F)$ is called the *test power under the alternative F.* This value characterizes the probability of making a correct decision (rejecting H_0) when H_0 is not true. A test whose power under the alternatives is greater is chosen as the best one compared to other tests with the significance level α.

Unbiasedness is a desirable property for the $\mathscr{T}_{1\alpha}$-test. This means that the condition

$$W(\mathscr{T}_{1\alpha};\, F) \leqslant \alpha \quad \forall F \in H_0 \tag{3.3}$$

must be met in addition to the condition

$$W(\mathscr{T}_{1\alpha};\, F) \geqslant \alpha \quad \forall F \in H_1 \tag{3.4}$$

(i.e., the probability of getting into the critical region must be greater under the alternative than under the null hypothesis).

The power function cannot always be found (for this we must know the distribution of the test statistic under all the alternatives), but it is frequently possible to investigate its asymptotic behaviour when the sample size n tends to infinity (to show that the power function depends on the size of a sample, we write $W_n(F)$). When studying the asymptotic properties of tests, we first check whether they are *consistent*. Consistency implies that

$$\lim_{n \to \infty} W_n(F) = 1 \quad \forall F \in H_1. \tag{3.5}$$

Thus, when the number of observations is large, a consistent test shows any deviations from the null hypothesis with a probability close to unity, i.e., if any fixed alternative is true and n is large, then we get into the critical region with a probability close to unity, and hence reject the null hypothesis which is false (i.e., we take a correct decision).

Other properties of a consistent test can be investigated when studying the asymptotic behaviour of the power $W_n(F_n)$ under "close" alternatives F_n, i.e., when the sequence $\{F_n\}$ of the alternatives gets closer (in some sense) to the null hypothesis H_0 as $n \to \infty$. The "threshold" case of finding a sequence $\{F_n\}$ for which

$$\lim_{n \to \infty} W_n(F_n) = \gamma, \quad \alpha < \gamma < 1, \tag{3.6}$$

is the most interesting. Here we have to calculate γ.

3.2. *Kolmogorov's test* and the χ^2-*test* are frequently used to verify the hypothesis H_0: $F_\xi(x) = F(x)$.

Kolmogorov's test is applied when $F(x)$ is a continuous function. The maximum deviation $D_n = D_n(\mathbf{X}) = \sup\limits_{-\infty < x < \infty} |F_n(x) - F(x)|$ of the empirical distribution function $F_n(x)$ from the hypothetical function $F(x)$ is the test statistic. For fixed x the value of $F_n(x)$ is an optimum estimator for $F(x)$ and, as n grows, we have $F_n(x) \to F(x)$. This means that when the hypothesis H_0 is true, D_n does not essentially deviate from zero at least for large n. The limiting Kolmogorov's distribution $K(t) = \sum\limits_{j=-\infty}^{\infty} (-1)^j \exp\{-2j^2t^2\}$ for which published tables are available gives a good approximation of the exact distribution $\mathbf{P}(\sqrt{n}D_n \leqslant t)$ for $n \geqslant 20$.

The critical region of the test is defined by the inequality $\sqrt{n}D_n \geqslant t_\alpha$, where $K(t_\alpha) = 1 - \alpha$. For example, $t_{0.1} = 1.23$, $t_{0.05} = 1.36$, $t_{0.01} = 1.63$.

The original statistical data are often grouped preliminarily. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be repeated independent observations on a random variable ξ with the set of possible values Δ. We consider a partition $\Delta = \Delta_1 \cup \ldots \cup \Delta_N$, $\Delta_i \cap \Delta_j = \varnothing$, $i \neq j$, and suppose that ν_j is the number of the units of the sample $\mathbf{X}$ in the subset Δ_j, and $p_j = p_j(F) = \int_{\Delta_j} dF(x)$ is the probability that ν_j is in Δ_j for the given distribution F of ξ, $j = 1, \ldots, N$ ($\nu_1 + \ldots + \nu_N = n$, $p_1 + \ldots + p_N = 1$). Then the frequency vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ has a polynomial distribution $M(n; \mathbf{p} = (p_1, \ldots, p_N))$ and every hypothesis on the distribution $\mathscr{L}(\xi)$ is transformed into a respective hypothesis about the vector $\mathbf{p}$ from the distribution $M(n; \mathbf{p})$. Thus, the given method implies a transition from the original observations $\mathbf{X} = (X_1, \ldots, X_n)$ to the frequencies $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ with which the sample units get in the respective subsets $\Delta_1, \ldots, \Delta_N$. Thus representation of statistical data is called the *method of grouped observations*, and the subsets

$\Delta_1, \ldots, \Delta_N$ are called the *grouping intervals*. The relative frequency ν_j/n of getting into the interval Δ_j is a consistent estimate for the probability p_j, and we may choose various functions of the differences $\left|\frac{\nu_j}{n} - p_j^0\right|$, $j = 1, \ldots, N$, as a measure of discrepancy between the empirical data and the hypothetical values $\mathbf{p}^0$. The measure

$$T = X_n^2 = \sum_{j=1}^{N} \frac{(\nu_j - np_j^0)^2}{np_j^0}$$

suggested by K. Pearson is frequently used. If H_0 is a simple hypothesis which uniquely describes the probabilities $\mathbf{p}^0 = (p_1^0, \ldots, p_N^0)$, then for $0 < p_j^0 < 1$, $j = 1, \ldots, N$, and $n \to \infty$ the respective goodness of fit test (called the χ^2-*test*) is asymptotically defined by the critical region $\{X_n^2 \geqslant \chi^2_{1-\alpha, N-1}\}$, where $\chi^2_{p,r}$ is a p-quantile of the distribution $\chi^2(r)$. Other applications of this method can be found in [7, Chap. 3].

3.3. Verifying the homogeneity of statistical data is an important problem. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)$ be independent samples which describe the same process or phenomenon but are generally obtained under different conditions. We have to find out whether they are samples from the same distribution or whether the distribution law has changed from sample to sample, i.e., we have to verify the homogeneity hypothesis H_0 that $F_1(x) = F_2(x)$, where $F_1(x)$ and $F_2(x)$ are the distribution functions of $\mathbf{X}$ and $\mathbf{Y}$, respectively. *Smirnov's homogeneity test* for continuous distributions is frequently applied in this case. The test is based on the statistic $D_{nm} = D_{nm}(\mathbf{X}, \mathbf{Y}) = \sup_{-\infty < x < \infty} |F_{1n}(x) - F_{2m}(x)|$, where $F_{1n}(x)$ and $F_{2m}(x)$ are the empirical distribution functions constructed from the samples $\mathbf{X}$ and $\mathbf{Y}$, respectively. When the hypothesis H_0 is true, the functions $F_{1n}(x)$ and $F_{2m}(x)$ get closer as the sizes n and m of the samples grow, and therefore the statistic D_{nm} practically does not deviate from zero. The exact distribution of $\mathbf{P}\left(\sqrt{\frac{nm}{n+m}}\, D_{nm} \leqslant t\right)$ is approximated by the limiting Kolmogorov's distribution $K(t)$. The critical region of the test is found from the inequality $\sqrt{\frac{nm}{n+m}}\, D_{nm} \geqslant t_\alpha$, where $K(t_\alpha) = 1 - \alpha$.

The *homogeneity* χ^2-*test* is frequently used to verify the homogeneity of discrete data or the data which can be made discrete by the method of grouped observations. This method is also used to compare any number of samples.

Let us carry out k series of independent observations of sizes $n_1, \ldots, n_k$ and observe in each series a variable feature assuming one of the s possible values (outcomes). Let ν_{ij} be the number of the realizations of the ith outcome in the jth series $\left(\sum_{i=1}^{s} \nu_{ij} = n_j,\right.$ $\left. j = 1, \ldots, k\right)$. We will test the hypothesis H_0 that all the observations were carried out on the same random variable. The quantity

$$X_n^2 = n\left(\sum_{i=1}^{s}\sum_{j=1}^{k} \frac{\nu_{ij}^2}{n_j \nu_{i\cdot}} - 1\right),$$

where $\nu_{i\cdot} = \sum_{j=1}^{k} \nu_{ij}$, $i = 1, \ldots, s$, $n = n_1 + \ldots + n_k$, is in our case the statistic of the χ^2-test.

The critical region is defined as $X_n^2 \geqslant \chi^2_{1-\alpha,(s-1)(k-1)}$, where the test boundary is found in the tables for the quantiles of the χ^2-distribution. If n is sufficiently large, the probability that the true hypothesis will be rejected is approximately α.

3.4. If $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample from the distribution $\mathscr{L}(\xi)$ and the set $\mathscr{F}$ of all the admissible distributions of the observable random variable ξ is given in a parametric form as $\mathscr{F} = \{F(x; \theta), \theta = (\theta_1, \ldots, \theta_r) \in \Theta\}$, then the hypotheses about the distribution $\mathscr{L}(\xi)$ are formulated in terms of the unknown parameter θ and are called *parametric*. In the general case the parametric (null) hypothesis is given in the form H_0: $\theta \in \Theta_0$ for a subset $\Theta_0 \subset \Theta$. Then the alternative hypothesis is of the form H_1: $\theta \in \Theta_1 = \Theta \setminus \Theta_0$. Thus, in a parametric model the alternative hypothesis has a form similar to that of the null hypothesis, and a deviation from the null hypothesis is equivalent to accepting a concrete alternative.

In the general theory of testing parametric hypotheses the tests are directly specified by the respective critical regions in the sample space $\mathscr{X} = \{\mathbf{x} = (x_1, \ldots, x_n)\}$. Thus, the test for verifying H_0 at a significance level α is given by a subset $\mathscr{X}_{1\alpha} \subset \mathscr{X}$ for which the condition

$$\mathbf{P}_\theta(\mathscr{X} \in \mathscr{X}_{1\alpha}) \leqslant \alpha \quad \forall \theta \in \Theta_0 \tag{3.7}$$

(an analogue of (3.1)) is met. Then the test (called the $\mathscr{X}_{1\alpha}$-*test*) is constructed as follows. If $\mathbf{x}$ is the observed realization of the sample $\mathbf{X}$, then the hypothesis H_0 is rejected for $\mathbf{x} \in \mathscr{X}_{1\alpha}$ (the alternative hypothesis H_1 is accepted), and if $\mathbf{x} \in \mathscr{X}_{0\alpha} = \overline{\mathscr{X}_{1\alpha}}$, then the hypothesis H_0 is

accepted. The power function is in this case written as

$$W(\theta) = W(\mathscr{X}_{1\alpha};\ \theta) = \mathbf{P}_\theta(\mathbf{X} \in \mathscr{X}_{1\alpha}), \quad \theta \in \Theta$$

(compare it with (3.2)).

The probabilities of erroneous decisions for the $\mathscr{X}_{1\alpha}$-test are expressed through its power function as follows. The *probability of Type I error* (rejecting H_0 when it is true) is equal to $W(\theta)$, $\theta \in \Theta_0$ (in symbols we write $\mathbf{P}(H_1|H_0)$), and the *probability of Type II error* (accepting H_0 when it is false) is equal to $1 - W(\theta)$, $\theta \in \Theta_1$ (in symbols we write $\mathbf{P}(H_0|H_1)$).

We now formulate a rational principle for choosing the critical region in terms of the probabilities of errors, i.e., *for a given number of trials we find the boundary for the probability of Type I error choosing the critical region for which the probability of Type II error is minimal.*

Let $\mathscr{X}_{1\alpha}$ and $\mathscr{X}^*_{1\alpha}$ be two tests of the same significance level α for the hypothesis H_0. If $W(\mathscr{X}^*_{1\alpha};\ \theta) \leqslant W(\mathscr{X}_{1\alpha};\ \theta)$ for $\theta \in \Theta_0$ and $W(\mathscr{X}^*_{1\alpha};\ \theta) \geqslant W(\mathscr{X}_{1\alpha};\ \theta)$ for $\theta \in \Theta_1$ (the strict inequality holding for at least one $\theta \in \Theta_1$), then we say that the $\mathscr{X}^*_{1\alpha}$-test is *uniformly more powerful* compared to the $\mathscr{X}_{1\alpha}$-test, the first test being preferable because it leads to smaller errors. If the indicated inequalities hold for any $\mathscr{X}_{1\alpha}$, then $\mathscr{X}^*_{1\alpha}$ is called a *uniformly most powerful* (u.m.p.) *test*. If H_1 is a simple alternative (the set Θ_1 consists of one point), then we use a *most powerful test* instead of a uniformly most powerful test. In some cases this method of comparing tests allows us to find an optimal (best) test for a given problem. It is sometimes possible to construct optimal tests in the class of *unbiased tests*, i.e., when the condition $W(\theta) \geqslant \alpha$ $\forall\theta \in \Theta_1$ is met in addition to (3.7).

Theoretically, it is sometimes convenient to deal with the *randomized tests* when for a given observation $\mathbf{x}$ the hypothesis H_0 is rejected with the probability $\varphi(\mathbf{x})$ and accepted with the complementary probability $1 - \varphi(\mathbf{x})$. The function $\varphi(\mathbf{x})$, $0 \leqslant \varphi(\mathbf{x}) \leqslant 1$, $\mathbf{x} \in \mathscr{X}$, is called a *critical function*. The construction of the non-randomized $\mathscr{X}_{1\alpha}$-test described above corresponds to the case when $\varphi(\mathbf{x})$ is an indicator of the set $\mathscr{X}_{1\alpha}$, i.e., $\varphi(\mathbf{x}) = 1$ for $\mathbf{x} \in \mathscr{X}_{1\alpha}$ and $\varphi(\mathbf{x}) = 0$ for $\mathbf{x} \notin \mathscr{X}_{1\alpha}$. The power function of a randomized test is defined by the relation $W(\theta) = W(\varphi;\ \theta) = \mathbf{E}_\theta\varphi(\mathbf{X})$.

3.5. Most of the methods for constructing optimal tests are based on the Neyman-Pearson theory of tests which states that verifying a simple hypothesis against a simple alternative one can find a most powerful test. Indeed, if $\Theta = \{\theta_0, \theta_1\}$, then for any significance level α the most powerful test for verifying the hypothesis H_0: $\theta = \theta_0$

against the alternative H_1: $\theta = \theta_1$ exists and is defined by the critical region

$$\mathscr{X}^*_{1\alpha} = \left\{ \mathbf{x};\ l(\mathbf{x}) = \frac{L(\mathbf{x};\ \theta_1)}{L(\mathbf{x};\ \theta_0)} \geqslant c_\alpha \right\}, \tag{3.8}$$

where $L(\mathbf{x};\ \theta) = \prod_{i=1}^{n} f(x_i;\ \theta)$ is the likelihood function (see [7, Sec. 4.2] on some properties of discrete observations).

If we test a simple hypothesis H_0: $\theta = \theta_0$ against a composite alternative H_1: $\theta \in \Theta \setminus \{\theta_0\}$, then a uniformly most powerful test exists when the critical region $\mathscr{X}^*_{1\alpha} = \mathscr{X}^*_{1\alpha}(\theta_0;\ \theta_1)$ defined in (3.8) is independent of the concrete $\theta_1 \in \Theta \setminus \{\theta_0\}$. In this case $\mathscr{X}^*_{1\alpha}$ is a u.m.p. test. This is typical for an important class of models $\mathscr{F}$ with a *monotone likelihood ratio* (i.e., the models having a sufficient statistic $T(\mathbf{X})$, where the function $l(\mathbf{x}) = g(T(\mathbf{x});\ \theta_1)/g(T(\mathbf{x});\ \theta_0)$ is monotone in T (see the factorization test in Sec. 2.3)), and also for one-sided alternatives $H_1^\pm$: $\theta \gtrless \theta_0$ (θ is a scalar) [7, p. 192]. Moreover, for such models the u.m.p. test for verifying a simple hypothesis H_0: $\theta = \theta_0$ against a right-sided alternative H_1^+: $\theta > \theta_0$ is simultaneously a u.m.p. test for verifying a composite hypothesis H_0: $\theta \leqslant \theta_0$ against H_1^+ of the same significance level (a similar statement is also true for a dual problem of testing H_0: $\theta \geqslant \theta_0$ against H_1^-: $\theta < \theta_0$ [10]).

Specifically, for the exponential model defined by the density

$$f(x;\ \theta) = \exp\{A(\theta)B(x) + C(\theta) + D(x)\},$$

the statistic $T(\mathbf{X}) = \sum_{i=1}^{n} B(X_i)$ is sufficient, and if the function $A(\theta)$ is strictly monotone, the u.m.p. $\mathscr{X}^*_{1\alpha}$-tests have the form given in Table 3.1.

Table 3.1

	H_1^+: $\theta > \theta_0$	H_1^-: $\theta < \theta_0$
$A(\theta)\uparrow$	$\{T(\mathbf{x}) \geqslant c_\alpha^+\}$	$\{T(\mathbf{x}) \leqslant c_\alpha^-\}$
$A(\theta)\downarrow$	$\{T(\mathbf{x}) \leqslant c_\alpha^+\}$	$\{T(\mathbf{x}) \geqslant c_\alpha^-\}$

In some cases when testing a simple hypothesis H_0: $\theta = \theta_0$ against a two-sided alternative H_1: $\theta \neq \theta_0$, it is also possible to construct a u.m.p. unbiased test [7, p. 196].

It is sometimes possible to find the solution of this problem using the following technique. If the $\mathscr{X}^-_{1\alpha_1}$ and $\mathscr{X}^+_{1\alpha_2}$ u.m.p. tests exist in the investigated model against the one-sided alternatives H_1^- and H_1^+, respectively, then a test of the form $\mathscr{X}_{1\alpha} = \mathscr{X}^-_{1\alpha_1} \cup \mathscr{X}^+_{1\alpha_2}$ is used, where $\alpha_1 + \alpha_2 = \alpha$.

The case of small deviations from the null hypothesis H_0: $\theta = \theta_0$ is especially interesting. When investigating the properties of a test, we may restrict ourselves to the analysis of the local behaviour of the power function $W(\theta)$ in the neighbourhood of the point θ_0. This approach allows us to construct a *local most powerful test* even if the u.m.p. test does not exist [7, p. 199].

3.6. In many cases the fact that testing a simple hypothesis about θ is an inverse problem to that of constructing a confidence set for θ can help greatly. Indeed, if $\mathscr{G}_\gamma(\mathbf{X})$ is a γ-confidence set for θ, then $\mathscr{X}_{0\alpha} = \{\mathbf{x}: \theta_0 \in \mathscr{G}_\gamma(\mathbf{x})\}$ defines the acceptance region for the hypothesis H_0: $\theta = \theta_0$ with the significance level $\alpha = 1 - \gamma$. The converse is also true, i.e., if for every $\theta_0 \in \Theta$ there is a test $\mathscr{X}_{1\alpha} = \mathscr{X}_{1\alpha}(\theta_0)$ for verifying the hypothesis H_0: $\theta = \theta_0$, then, having found the subset $\mathscr{G}_\gamma(\mathbf{x}) = \{\theta: \mathbf{x} \in \overline{\mathscr{X}_{1\alpha}(\theta)}\}$, $\gamma = 1 - \alpha$, for every $\mathbf{x} \in \mathscr{X}$, we will prove that $\mathscr{G}_\gamma(\mathbf{X})$ is a γ-confidence set for θ. Thus, if one of these problems is solved for a certain model, then this algorithm can be used to solve the other problem. Here the u.m.p. tests correspond to the shortest confidence sets and vice versa.

3.7. The *likelihood ratio method* (l.r.m.) is universal for constructing the tests for verifying composite parametric hypotheses. The general form of the *likelihood ratio test* for testing a hypothesis H_0: $\theta \in \Theta_0$ is

$$\mathscr{X}_{1\alpha} = \mathscr{X}_{1\alpha}(\Theta_0, \Theta) = \{\mathbf{x}: \lambda_n(\mathbf{x}) = \sup_{\theta \in \Theta_0} L(\mathbf{x}; \theta) / \sup_{\theta \in \Theta} L(\mathbf{x}; \theta) \leqslant c_\alpha\},$$

where the boundary c_α is chosen from the condition

$$W(\theta) = \mathbf{P}_\theta(\lambda_n(\mathbf{X}) \leqslant c_\alpha) \leqslant \alpha \quad \forall \theta \in \Theta_0.$$

In practice this approach gives satisfactory results. Besides under some conditions the likelihood ratio test possesses the optimality for large samples.

If the regularity conditions which ensure the existence, uniqueness, and asymptotic normality of the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{1n}, \ldots, \hat{\theta}_{rn})$ for the parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$ are met (see Sec. 2.4), then, given a simple hypothesis H_0: $\theta = \theta_0$ for large samples, the likelihood ratio test is asymptotically defined by the critical region

$$\mathscr{X}_{1\alpha} = \{\mathbf{x}: -2 \ln \lambda_n(\mathbf{x}) \geqslant \chi^2_{1-\alpha, r}\},$$

where $\chi^2_{p,r}$ is a p-quantile of the distribution $\chi^2(r)$. This test is consistent ($W_n(\theta) \to 1$ for $n \to \infty$ $\forall\theta \neq \theta_0$), and as $n \to \infty$ its density for the close alternatives of the form $\theta_1^{(n)} = \theta_0 + \beta/\sqrt{n}$, $\beta = (\beta_1, \ldots, \beta_r) \neq 0$, satisfies the relation

$$W_n(\theta^{(n)}) \to 1 - F_r(\chi^2_{1\alpha,r}; \lambda^2),$$

where $\lambda^2 = \beta' \mathbf{I}(\theta_0)\beta$, $\mathbf{I}(\theta)$ is the information matrix of the model, and $F_r(t; \lambda^2)$ is a function of the non-central χ^2-distribution with r degrees of freedom and the skewness parameter λ^2 [7, p. 210]. The likelihood ratio test possesses similar asymptotic properties when the hypothesis H_0 is composite [7, pp. 211-213].

Problems

Goodness of Fit Tests

3.1. Given the data of Problem 1.13, check whether they are consistent with the hypothesis H_0 that the coin was symmetric. Take the significance level (a) 0.05; (b) 0.1.

3.2. Given the data of Problem 1.14, test the hypothesis H_0 that the numbers are random. For what significance level should the hypothesis H_0 be rejected?

3.3. In $n = 4000$ independent trials the events A_1, A_2, A_3 which constitute a complete group were realized 1905, 1015, 1080 times, respectively. Given the significance level 0.05, check whether these data are consistent with the hypothesis H_0: $p_1 = 1/2$, $p_2 = p_3 = 1/4$, where $p_i = \mathbf{P}(A_i)$.

3.4. The number π written in decimal form contains in the first 10 002 positions after the decimal point the digits 0, 1, ..., 9 respectively 968, 1026, 1021, 974, 1014, 1046, 1021, 970, 948, 1014 times [1]. Given the significance level 0.05, can we consider these digits to be random numbers? For what significance level should the hypothesis be rejected?

3.5. Are the data in Problems 1.16 and 1.17 consistent with the hypothesis that the dice are symmetric?

3.6. A large batch of goods may contain some defective items. The supplier assumes that they constitute 3% of the batch, while the buyer insists on 10%. The contract stipulates that if 20 randomly chosen items contain no more than one defective item, the batch is accepted on the supplier's terms, otherwise it is accepted on the terms of the buyer. Find (1) the statistical hypotheses, the test statistic, its domain,

and critical region; (2) the distribution of the test statistic, Type I and Type II errors, and their probabilities.

3.7. Given the significance level 0.1, are the data in Problem 1.19 consistent with the hypothesis H_0 that the times shown by the watches are uniformly distributed on the interval (0, 12)? For what significance levels is the hypothesis H_0 accepted?

3.8. When breeding pea-plants, Mendel observed the frequencies of various seeds produced by hybrids of round yellow peas and wrinkled green peas. These data and the respective probabilities as predicted by Mendel's theory of heredity are given in the following table:

Seeds	Frequency	Probability
Round and yellow	315	9/16
Wrinkled and yellow	101	3/16
Round and green	108	3/16
Wrinkled and green	32	1/16
Total	$n = 556$	1

Check the hypothesis H_0 that the frequency data are consistent with the theoretical probabilities (at the significance level $\alpha \leqslant 0.9$).

3.9. Using a table for some function ($\cos x$, e^x, $\ln x$, etc.), write out 100 digits choosing the second digit after the point at each value. Test the hypothesis that the numbers 0, 1, ..., 9 are random at the significance level (a) 0.05; (b) 0.01.

3.10. Grouping the data in Problem 1.21 into $N = 4$ equiprobable intervals (under the hypothesis H_0), test the hypothesis H_0: $F_\xi(x) = 1 - e^{-x}$, $x \geqslant 0$ (at the significance level 0.1).

3.11. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$, test the hypothesis that the observable random variable ξ is distributed exponentially, i.e., H_0: $F_\xi(x) = 1 - e^{-x/\theta}$, $x \geqslant 0$ (the parameter $\theta > 0$ is unknown). Applying the method of grouped observations with the intervals $\Delta_j = [(j - 1)a, ja)$, $j = 1, \ldots, N - 1$, $\Delta_N = [(N - 1)a, \infty)$, where $a > 0$ is a given number, construct a χ^2 goodness of fit test for the hypothesis H_0. Analyze the data in Problem 1.21 from this point of view, assuming that $N = 3$, $a = 1$.

3.12. In Fisher's genetic model [1] the probabilities of four types of the offspring are

$$p_1(\theta) = \frac{2 + \theta}{4}, \quad p_2(\theta) = p_3(\theta) = \frac{1 - \theta}{4}, \quad p_4(\theta) = \frac{\theta}{4},$$

where $\theta \in (0, 1)$ is an unknown parameter. Construct the χ^2-test to check whether this model is consistent with the actual data.

3.13. In 8000 independent trials the events A, B, C which form a complete group occurred 2014, 5012, 974 times, respectively. Is the hypothesis H_0: $\mathbf{P}(A) = 0.5 - 2\theta$, $\mathbf{P}(B) = 0.5 + \theta$, $\mathbf{P}(C) = \theta$, $0 < \theta < 0.25$, true at the significance level 0.05?

Hint. Use the solution to Problem 3.12.

3.14. Test the hypothesis H_0: $\mathscr{L}(\xi) = \Pi(\theta)$, where θ is an unknown parameter, for the data of Problem 1.23.

Hint. Take the sample mean as the estimate for the unknown parameter θ [7, p. 152].

3.15. The number of gold particles ξ in a thin layer of suspension under a microscope was registered in equal time intervals. Using the data from the table

The number of particles	0	1	2	3	4	5	6	7	Total
m_i	112	168	130	68	32	5	1	1	$\Sigma m_i = 518$

test the hypothesis H_0: $\mathscr{L}(\xi) = \Pi(\theta)$, where θ is an unknown parameter.

3.16. The table below gives the number m_i of 0.25-km^2 plots in the southern part of London each of which has been bombed i times during the World War II. Check whether these data are consistent with the Poisson distribution law at the significance level $\alpha = 0.05$.

i	0	1	2	3	4	5 and more	Total
m_i	229	211	93	35	7	1	$\Sigma m_i = 576$

3.17. Of 2020 families with two children, 527 families have two boys, 476 families have two girls, and the remaining 1017 families have children of both sexes. Given the significance level 0.05, can we consider that the number of boys in two-children families is a binomial random variable?

3.18. Among the 2000 individuals, 181 suffered from flu once, 9 had it twice, and the remaining 1810 were healthy. Given the significance level 0.05, do these data correspond to the hypothesis that the number of illnesses an individual suffers is a binomial random variable?

Hint. See the solution to Problem 3.17.

3.19*. Investigate the asymptotic behaviour (as $n \to \infty$) of the mean and variance of the statistic X_n^2 of the χ^2-test against "close" alternatives of the form

$$H_1^{(n)}: \ p_j = p_j^{(n)} = p_j^0 + \frac{\beta_j}{\sqrt{n}}, \ \ j = 1, \ldots, N, \quad \sum_{j=1}^{N} \beta_j = 0.$$

Hint. Use the formulas for $\mathbf{E}(X_n^2|\mathbf{p})$ and $\mathbf{D}(X_n^2|\mathbf{p})$ given in [7, p. 145].

3.20*. Let $\mu_0 = \mu_0(n, N)$ be the number of empty intervals when n observations are grouped into N equiprobable (under the hypothesis H_0) intervals. Consider the hypotheses of the form

$$H_1^{(n)}: \ p_j = p_j^{(n)} = \frac{1}{N}\left(1 + \frac{b_j}{n^{1/4}}\right), \ \ j = 1, \ldots, N,$$

where

$$\max_{1 \leqslant j \leqslant N} |b_j| \leqslant c < \infty, \ \sum_{j=1}^{N} b_j = 0, \ b^2(N) \equiv \frac{1}{N}\sum_{j=1}^{N} b_j^2 \to b^2 > 0$$

for $N \to \infty$.

Prove that for $n, N \to \infty$, $n/N = \varrho > 0$ we have

$$\mathbf{E}(\mu_0|H_1^{(n)}) = Ne^{-\varrho} + \frac{\sqrt{N}}{2} b^2(N)\varrho^{3/2}e^{-\varrho} + O(N^{1/4}),$$

$$\mathbf{D}(\mu_0|H_1^{(n)}) = Ne^{-\varrho}(1 - e^{-\varrho}(1 + \varrho))(1 + O(N^{-1/2})).$$

Hint. Use formulas (3.16) from [7, p. 155].

3.21. Entrants to a university are divided into two groups, 300 people in each group. In the first group 33, 43, 80, 144 individuals got grades "2", "3", "4", "5", respectively. The data for the second group were 39, 35, 72, 154 individuals, respectively. Can we consider both groups to be homogeneous at the significance level 0.05?

3.22. The table

n_j	1072	1133	2455	1995
ν_j	22	23	49	33

gives the data on the death rate of mothers having their first baby in four periods of time [1] (n_j is the number of mothers, ν_j is the number of deaths). Test the hypothesis H_0 that the death rates in these periods do not differ.

| *Hint.* Use the homogeneity χ^2-test for trials with two outcomes.

3.23*. Let two series of n_1 and n_2 independent trials be carried out, each of which has either the outcome A or the outcome $\overline{A}$. The results are given in the table

	(1)	(2)	Σ
A	ν_{11}	ν_{12}	$\nu_{1\cdot}$
$\overline{A}$	ν_{21}	ν_{22}	$\nu_{2\cdot}$
Σ	$\nu_{\cdot 1} = n_1$	$\nu_{\cdot 2} = n_2$	$n = n_1 + n_2$

where the columns show the number of the realizations of the respective outcomes in each series.

(1) Make sure that the statistic X_n^2 used to test the hypothesis H_0 that the trials are homogeneous can be written as $X_n^2 = Z_n^2$, where

$$Z_n = \left(\frac{\nu_{11}}{n_1} - \frac{\nu_{12}}{n_2}\right)\sqrt{\frac{nn_1n_2}{\nu_{1\cdot}\nu_{2\cdot}}}.$$

(2) Prove that $\mathscr{L}(Z_n|H_0) \to \mathscr{N}(0, 1)$ as $n_1, n_2 \to \infty$ and construct a test for verifying the hypothesis H_0: $p_1 = p_2$ against the one-sided alternative H_1: $p_1 > p_2$ (here p_i is the probability of the realization A in the trials of the ith series, $i = 1, 2$).

3.24. Let $\nu_1, \ldots, \nu_N$ be independent random variables with $\mathscr{L}(\nu_i) = \Pi(\theta_i)$, $i = 1, \ldots, N$, where the parameters θ_i are unknown. Assume that $\nu_1 + \ldots + \nu_N = n$ and construct a test for verifying the homogeneity hypothesis H_0: $\theta_1 = \ldots = \theta_N$.

| *Hint.* Use Problem 1.54.

3.25*. (*The empty box test.*) Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ is a sample from the distribution $\mathscr{L}(\xi) = R(0, 1)$, $0 = X_{(0)} \leqslant X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)} \leqslant X_{(n+1)} = 1$ is its ordered sample, and $B_i = (X_{(i-1)}, X_{(i)}]$, $i = 1, \ldots, n + 1$, are the *sampling blocks* generated by it. Suppose that $\mathbf{Y} = (Y_1, \ldots, Y_m)$ is a sample from a different distribution $\mathscr{L}(\eta)$ on the interval $[0, 1]$ which is independent of $\mathbf{X}$. The distribution function $F(x)$ of $\mathscr{L}(\eta)$ has the density $f(x) = F'(x)$. The number of units of the sample $\mathbf{Y}$, which are in the block B_i, $i = 1, \ldots, n + 1$, is denoted $\varkappa_i = \varkappa_i(n, m)$.

(1) Prove that under the homogeneity hypothesis H_0: $\mathscr{L}(\xi) = \mathscr{L}(\eta)$ the vector of the block frequencies $\varkappa = (\varkappa_1, \ldots, \varkappa_{n+1})$ assumes all the possible values with the same probability $(C^n_{n+m})^{-1}$. Show that the conditional distribution $\mathscr{L}(\xi_1, \ldots, \xi_{n+1} | \xi_1 + \ldots + \xi_{n+1} = m)$, where the random variables $\xi_1, \ldots, \xi_{n+1}$ are independent and have the geometric distribution $\bar{B}i(1, p)$ with an arbitrary $p \in (0, 1)$, has the same form.

(2) Consider the statistic $s_0(n, m)$ (the number of the empty blocks), viz.,

$$s_0(n, m) = \sum_{i=1}^{n+1} I(\varkappa_i = 0),$$

where $I(\cdot)$ is an indicator, and use the representation

$$\mathscr{L}(s_0(n, m)) = \mathscr{L}\left(\sum_{i=1}^{n+1} I(\xi_i = 0) | \xi_1 + \ldots + \xi_{n+1} = m \right),$$

which stems from Sec. 3.1, to prove that $s_0(n, m)$ has a hypergeometric distribution $H(n + 1, n + m, n)$. Derive from this the expression for the mean and variance of the statistic $s_0(n, m)$ under the hypothesis H_0.

(3) Prove that if $n, m \to \infty$ so that $m/n = \varrho > 0$, then

$$\mathscr{L}(s_0(n, m) | H_0) \sim \mathscr{N}(n/(1 + \varrho), n\varrho^2/(1 + \varrho)^3).$$

(4) Prove that under the specified conditions for any alternative H_1 defined by the density $f(x) \not\equiv 1$, $x \in [0, 1]$, we will have

$$\mathbf{E}\left(\frac{s_0(n, m)}{n + 1} \middle| H_1\right) \to \int_0^1 \frac{dx}{1 + \varrho f(x)} > \frac{1}{1 + \varrho}.$$

Using these results, formulate the empty box test for verifying the homogeneity hypothesis H_0 [7, p. 162].

Hints. (1) Use the fact that the conditional distribution of the vector $\varkappa = (x_1, \ldots, x_{n+1})$ for the fixed values $(X_{(1)}, \ldots, X_{(n)}) = (x_1, \ldots, x_n)$ is a polynomial distribution $M(m; x_1, x_2 - x_1, \ldots, x_n - x_{n-1}, 1 - x_n)$. Then use Problems 1.39 (5) and 1.31.

(2) Consider the independent random variables $\bar{\xi}_i$ distributed as $\mathbf{P}(\bar{\xi}_i = r) = \mathbf{P}(\xi_i = r \mid \xi_i > 0)$, $r = 1, 2, \ldots$, $i = 1, 2, \ldots$, and show that

$$\mathbf{P}(\bar{\xi}_1 + \ldots + \bar{\xi}_s = m) = C_{m-1}^{s-1} q^s p^{m-s}, \quad q = 1 - p.$$

(3) Use the normal approximation for a binomial distribution, i.e., for $n \to \infty$ and $0 < p < 1$, $k = np + t\sqrt{npq}$, $|t| \leqslant c < \infty$,

$$b(k; n, p) = C_n^k p^k q^{n-k} = \frac{1 + o(1)}{\sqrt{2\pi npq}} e^{-t^2/2}.$$

Write the probability $\mathbf{P}(s_0(n, m) = k)$ in the form

$$\mathbf{P}(s_0(n, m) = k) = b(k; n + 1, p)$$
$$\times\, b(n - k; m - 1, p)/b(n; n + m, p), \quad p = 1/(1 + \varrho).$$

(4) Use Problem 1.31 to calculate $\mathbf{E}[I(x_l = 0) \mid H_1] = \mathbf{P}(x_l = 0 \mid H_1)$. While estimating the integral, apply the Cauchy-Schwarz inequality

$$\left(\int_0^1 g_1(x) g_2(x)\, dx\right)^2 \leqslant \int_0^1 g_1^2(x)\, dx \int_0^1 g_2^2(x)\, dx$$

for $g_1(x) = \sqrt{1 + \varrho f(x)}$, $g_2(x) = g_1^{-1}(x)$.

3.26. Test the independence hypothesis for the following bivariate contingency table:

ξ_1	ξ_2			Σ
	b_1	b_2	b_3	
a_1	3009	2832	3008	8849
a_2	3047	3051	2997	9095
a_3	2974	3038	3018	9030
Σ	9030	8921	9023	26 974

The significance level is 0.05.

3.27. Of 300 university entrants who passed the entrance examination, 97 obtained the top grade at school, and 48 got the top grade in the entrance examination, but only 18 obtained the top grade at school and in the examination. Test the hypothesis that the school marks and entrance marks are independent (the significance level is 0.1).

3.28*. Consider the following bivariate contingency table:

ξ_1	ξ_2: 1	ξ_2: 0	Σ
1	ν_{11}	ν_{12}	$\nu_{1\cdot}$
0	ν_{21}	ν_{22}	$\nu_{2\cdot}$
Σ	$\nu_{\cdot 1}$	$\nu_{\cdot 2}$	n

(1) Make sure that the statistic $\hat{X}_n^2$ [7, p. 166] for testing the hypothesis H_0 on the independence of the factors ξ_1 and ξ_2 can be represented as $\hat{X}_n^2 = Z_n^2$, where

$$Z_n = n^{3/2}\left(\nu_{11} - \frac{\nu_{1\cdot}\nu_{\cdot 1}}{n}\right)\Bigg/\sqrt{\nu_{1\cdot}\nu_{2\cdot}\nu_{\cdot 1}\nu_{\cdot 2}} = \left(\frac{\nu_{11}}{\nu_{\cdot 1}} - \frac{\nu_{12}}{\nu_{\cdot 2}}\right)\sqrt{\frac{n\nu_{\cdot 1}\nu_{\cdot 2}}{\nu_{1\cdot}\nu_{2\cdot}}}.$$

(2) Show that the sample correlation coefficient is $\varrho_n = Z_n/\sqrt{n}$ and hence $Z_n/\sqrt{n} \xrightarrow{P} \varrho = \operatorname{corr}(\xi_1, \xi_2)$ as $n \to \infty$ (see Problem 1.38). Derive the equations

$$\varrho = \frac{\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}} = [\mathbf{P}(A\,|\,B) - \mathbf{P}(A\,|\,\overline{B})]\left[\frac{\mathbf{P}(B)\mathbf{P}(\overline{B})}{\mathbf{P}(A)\mathbf{P}(\overline{A})}\right]^{1/2}$$

with the events $A = \{\xi_1 = 1\}$, $\overline{A} = \{\xi_1 = 0\}$, $B = \{\xi_2 = 1\}$, $\overline{B} = \{\xi_2 = 0\}$.

(3) Prove that $\mathscr{L}(Z_n | H_0) \to \mathscr{N}(0, 1)$ as $n \to \infty$ and then construct a test for verifying the hypothesis H_0 against the alternative H_1: $\mathbf{P}(A\,|\,B) > \mathbf{P}(A\,|\,\overline{B})$ which means that the events A and B are positively conjugate (the probability of the pair A and B is greater than that of the pair A and $\overline{B}$).

3.29. We have two groups of data classified by two features, i.e., "accepted A- rejected $\overline{A}$ for entrance to the college" and "male B - female $\overline{B}$":

	B	$\overline{B}$	Σ
A	97	40	137
$\overline{A}$	263	42	305
Σ	360	82	$n = 442$

	B	$\overline{B}$	Σ
A	235	38	273
$\overline{A}$	35	7	42
Σ	270	45	$n = 315$

For each table test the hypothesis H_0 that the features A and B are independent against the alternative H_1: $\mathbf{P}(A \mid B) > \mathbf{P}(A \mid \overline{B})$.

3.30. The table below [8] gives 818 cases classified according to two features, i.e., vaccinated against cholera A and healthy B

	B	$\overline{B}$	Σ
A	276	3	279
$\overline{A}$	473	66	539
Σ	749	69	818

Construct a test for verifying the hypothesis H_0 that the features A and B are independent against the alternative H_1 that A and B are positively conjugate (i.e., that the vaccination is effective).

3.31. Given the significance level 0.001, can we consider the sequence 1.05, 1.12, 1.37, 1.50, 1.51, 1.73, 1.85, 1.98 to be a realization of a random vector whose all eight components are independent similarly distributed random variables?

3.32*. Assuming that

$$\Phi_n(z) \equiv \mathbf{E}z^{T_n} = \frac{1}{n!} \prod_{r=1}^{n-1} (1 + z + \ldots + z^r)$$

is a representation of the generating function of the statistic T_n (the

number of inversions in a repeated random sample of size n) [7, p. 170], prove that $\mathscr{L}(T_n) \sim \mathcal{N}\left(\frac{n(n-1)}{4}, \frac{n^3}{36}\right)$ as $n \to \infty$.

Hint. Consider the characteristic function and show that

$$\mathbf{E} \exp \left\{ it \left(T_n - \frac{n(n-1)}{4} \right) \frac{6}{n^{3/2}} \right\} \to \exp \{-t^2/2\}$$

for $n \to \infty$ and $|t| \leqslant c < \infty$.

3.33. Test the hypothesis that the data in Problem 1.22 are random.

Hint. Use the asymptotic variant of the test based on the statistic T_n (see the previous problem).

3.34. Obtain samples (of sizes $n = 20, 50, 100$) of uniformly distributed random numbers. Use the χ^2- and Kolmogorov's test to verify the hypothesis that the distribution is uniform.

3.35. Obtain samples (of size $n = 100$) of approximately normally distributed numbers by summing up N uniformly distributed terms ($N = 4, 8, 12$). Use the χ^2- and Kolmogorov's tests to verify the hypothesis that the distribution is normal.

3.36. Obtain a sample X_i (of size $n = 200$) of uniformly distributed random numbers. Using Smirnov's test, verify that (X_{2i}, $i = 1, 2, \ldots, 100$) and (X_{2i+1}, $i = 0, 1, \ldots, 99$) are samples from the same distribution.

3.37. Simulate a sequence $\{X_i\}$ of polynomial pseudo-random variables assuming the values of $1, \ldots, N$. Form two samples (X_{2i}, $i = 1, \ldots, n$) and (X_{2i+1}, $i = 0, \ldots, n-1$) from this sequence and use the χ^2-test to verify the hypothesis that the values corresponding to these samples are independent. Work with $N = 2, 4, 10$, and $n = 100$.

3.38. Obtain samples (of sizes $n = 10, 20, 40, 100$) of uniform pseudo-random numbers. Using the statistics T_n (the number of inversions in the ordered series of the sample), test the hypothesis on randomness.

A Choice Between Two Simple Hypotheses

3.39. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the binomial distribution $Bi(k; \theta)$. Construct a Neyman-Pearson test to verify the hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta = \theta_1$, $0 < \theta_0 < \theta_1 < 1$, and calculate its power.

3.40. (Continued from Problem 3.39.) Show that for $n \to \infty$ the test can be asymptotically defined by the critical region

$$\{T \geqslant kn\theta_0 - u_\alpha \sqrt{kn\theta_0(1-\theta_0)}\}, \quad T = \sum_{i=1}^{n} X_i, \ \Phi(u_\alpha) = \alpha,$$

and its power $W_n(\theta_1)$ satisfies the relation

$$\lim_{n\to\infty} W_n(\theta_1^{(n)}) = \Phi\left(\beta\sqrt{\frac{k}{\theta_0(1-\theta_0)}} + u_\alpha\right)$$

for $\theta_1 = \theta_1^{(n)} = \theta_0 + \beta/\sqrt{n}$, $\beta > 0$.

Hint. Use the De Moivre-Laplace theorem.

3.41. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from Poisson's distribution $\Pi(\theta)$, construct a Neyman-Pearson test to verify the hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta = \theta_1$, $0 < \theta_0 < \theta_1$, and calculate its power. Investigate the asymptotic behaviour of the test's characteristics for large samples.

Hint. Use Problem 1.39 (4) and the normal approximation for Poisson's distribution with a growing parameter. Consider a "close" alternative of the form given in Problem 3.40.

3.42. Observe the number of successes before the first failure in an experiment used to verify the hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta = \theta_1$, $0 < \theta_0 < \theta_1 < 1$, in a Bernoulli scheme with an unknown probability of success θ. Construct the most powerful test at the significance level $\alpha = \theta_0^s$, where $s \geqslant 1$ is a given number, and show that the probability of Type II error for this test is $\beta = 1 - \theta_1^s$.

3.43. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the exponential distribution $\Gamma(\theta, 1)$. Construct a most powerful test to verify the hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta = \theta_1$ and calculate its power function.

Hint. Use the fact that $\mathscr{L}_\theta(2X_i/\theta) = \chi^2_{(2)}$ (see Problem 1.51) and the solution to Problem 1.39 (2).

3.44. Test the hypothesis H_0: $\theta = 0$ against the alternative H_1: $\theta = 1$ for Cauchy's distribution $C(\theta)$. Show that for the significance level $\alpha = 1/2 - (1/\pi)\arctan(1/2) \approx 0.352$ the most powerful test constructed from one observation has the form $\mathscr{X}^*_{1\alpha} = \{X \geqslant 1/2\}$ and its power is $1/2 + (1/\pi)\arctan(1/2) \approx 0.648$. If $\alpha = 1/\pi(\arctan 3 - \arctan 1) \approx 0.148$, then the test has the form $\mathscr{X}^*_{1\alpha} = \{1 \leqslant X \leqslant 3\}$, and its power is $(1/\pi)\arctan 2 \approx 0.352$.

3.45*. Construct a test to verify the hypothesis H_0: $\mathscr{L}(\xi) = R(-a, a)$ against the alternative H_1: $\mathscr{L}(\xi) = \mathscr{N}(0, \sigma^2)$ (the parameters a and σ are given) under the condition that the observable random variable ξ is distributed symmetrically about zero. Consider the case of a large sample. Analyze the data: -0.460, -0.114, -0.325, $+0.196$, -0.174 for $a = 1/2$ and $\sigma^2 = 0.09$.

Hint. Use the Central Limit Theorem.

3.46. In a sequence of independent trials the probabilities of positive outcomes are the same and equal to p. Construct a test to verify the

hypothesis H_0: $p = 0$ against the alternative H_1: $p = 0.01$ and find the smallest sample size for which the probabilities of Type I and Type II errors do not exceed 0.01.

3.47. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the distribution $\mathcal{N}(\theta, \sigma^2)$, find a most powerful test to distinguish between two simple hypotheses H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$. Calculate its power and show that it is unbiased.

3.48. (Continued from Problem 3.47.) Define the minimum sample size $n^* = n^*(\alpha, \beta)$ for which the probabilities of Type I and Type II errors are not greater than α and β, respectively.

3.49. Let $\overline{X}$ and $\overline{Y}$ be the sample means of two samples with sizes n and m from the distributions $\mathcal{N}(\theta_1, \sigma_1^2)$ and $\mathcal{N}(\theta_2, \sigma_2^2)$, respectively. Using the statistic $T = (\overline{X} - \overline{Y})/\sigma$, where $\sigma^2 = \sigma_1^2/n + \sigma_2^2/m$, construct a test to verify the hypothesis H_0: $\Delta = \theta_1 - \theta_2 = 0$ against the alternative H_1: $\Delta > 0$.

Suppose that the probabilities of Type I and Type II errors are α and β, respectively, and n is the size of the first sample. Find the minimum size m^* of the second sample, such that the probabilities of erroneous conclusions were not greater than α and β.

[*Hint.* Use the solution to Problems 3.47 and 3.48.

3.50. Given a sample of size n, construct a most powerful test to distinguish between two simple hypotheses with respect to an unknown variance of a normal distribution (the mean is known). Find the test's power.

3.51*. Given an observation X, distinguish between two distributions with the densities $f_0(x)$ (the hypothesis H_0) and $f_1(x)$ (the hypothesis H_1). Consider a test of the form

$$\mathscr{X}_1(c) = \{x: f_1(x) \geqslant cf_0(x)\}, \quad c > 0.$$

Let $\alpha(c)$ and $\beta(c)$ be the probabilities of Type I and Type II errors, respectively. Show that

(1) $\dfrac{\beta(c)}{1 - \alpha(c)} \leqslant c \leqslant \dfrac{1 - \beta(c)}{\alpha(c)}$;

(2) if $\alpha(c) + \beta(c) \leqslant 1$, the test is unbiased;

(3) $\min\limits_c (\alpha(c) + \beta(c)) = \alpha(1) + \beta(1)$ because the $\mathscr{X}_1(1)$-test minimizes the sum of the probabilities of the errors;

(4) suppose that $\mathbf{X}$ is a repeated sample of size n, i.e., $\mathbf{X} = (X_1, \ldots, X_n)$, $f_j(x) = \prod\limits_{i=1}^{n} f_j(x_i)$, $j = 0, 1$, and the probabilities of errors for the $\mathscr{X}_1(1)$-test are α_n and β_n. Prove that if $\int f_0(x) \ln (f_1(x)/$

$f_0(x))\,dx = \delta < 0$, then $\alpha_n, \beta_n \to 0$ as $n \to \infty$ (which means that we can completely distinguish between the hypotheses H_0 and H_1).

Hint. Write $\mathscr{X}_1(1) = \left\{x\colon\ T_n(x) \equiv \frac{1}{n} \sum_{i=1}^{n} \ln \frac{f_1(x_i)}{f_0(x_i)} \geqslant 0\right\}$ and apply the law of large numbers to the statistic $T_n(X)$. Note also that we always have $\delta \leqslant 0$ [7, p. 157] by Jensen's inequality.

3.52*. Let $\xi = (\xi_1, \ldots, \xi_r)$ be a normal random vector distributed as $\mathscr{N}(\mu^{(i)}, \mathbf{A})$, $i = 0, 1$, under the hypothesis H_i (the common covariance matrix $\mathbf{A}$ is supposed to be non-singular). Construct a Neyman-Pearson test to distinguish the hypothesis H_0 against the alternative H_1 by a single observation on ξ. Construct a test minimizing the sum of the probabilities of errors.

Composite Hypotheses

3.53. Given a binomial model $Bi(k, \theta)$, construct a uniformly most powerful test for verifying the hypothesis H_0: $\theta \leqslant \theta_0$ against the alternative H_1: $\theta > \theta_0$ for a sample of size n.

Hint. Use the property of a model with a monotone likelihood ratio and the solution to Problem 3.39.

3.54. Show that the Neyman-Pearson test constructed in Problem 3.41 (for Poisson's model $\Pi(\theta)$) is a u.m.p. test for verifying the hypothesis H_0: $\theta \leqslant \theta_0$ against the alternative H_1: $\theta > \theta_0$.

Hint. Use the solution to Problem 3.53.

3.55. Suppose that in a Bernoulli scheme the trials are carried out until the rth failure with an unknown probability of success θ, and T_r is the observed number of successes. Construct a u.m.p. test for verifying the hypothesis H_0: $\theta \leqslant \theta_0$ against the alternative H_1: $\theta > \theta_0$ and show that for $r \to \infty$ the respective critical boundary at the significance level α has the form $t_\alpha = (r\theta_0 - u_\alpha\sqrt{r\theta_0})/(1 - \theta_0)$, $\Phi(u_\alpha) = \alpha$.

Hint. Use the property of a model with a monotone likelihood ratio, the representation $T_r = X_1 + \ldots + X_r$, where $X_1, \ldots, X_r$ are independent and similarly distributed random variables, and $\mathscr{L}(X_1) = \overline{Bi}(1, \theta)$, and apply the Central Limit Theorem.

3.56. Show that the tests constructed in Problem 3.43 are u.m.p. tests for verifying the composite one-sided hypotheses H_0: $\theta \leqslant \theta_0$ against H_1: $\theta > \theta_0$ and H_0: $\theta \geqslant \theta_0$ against H_1: $\theta < \theta_0$.

3.57. (Sampling inspection.) Suppose that a batch of N items contains an unknown number θ of defective items, $\theta \in \{0, 1, \ldots, N\}$. In order to verify the hypothesis H_0: $\theta \leqslant \theta_0$ against the alternative

H_1: $\theta > \theta_0$, we test each of the n items chosen for control. Based on the statistic T (the number of defectives in the sample), construct a u.m.p. test.

Hint. Make sure that the distribution of T (the hypergeometric distribution $H(\theta, N, n)$) has a monotone likelihood ratio.

3.58. Given a normal model $\mathcal{N}(\theta, \sigma^2)$ with an unknown mean, construct a u.m.p. test for verifying the hypotheses H_0: $\theta \leqslant \theta_0$ against H_1: $\theta > \theta_0$ and H_0: $\theta \geqslant \theta_0$ against H_1: $\theta < \theta_0$.

Hint. Use the solution to Problem 3.47 and the properties of an exponential model.

3.59. Show that the test constructed in Problem 3.50 for the case $\theta_0 > \theta_1$ is a u.m.p. test for verifying a composite hypothesis H_0: $\theta \geqslant \theta_0$ against the left-sided alternative H_1: $\theta < \theta_0$ and, similarly, the test for $\theta_0 < \theta_1$ is a u.m.p. test for verifying the hypothesis H_0: $\theta \leqslant \theta_0$ against the right-sided alternative H_1: $\theta > \theta_0$.

Hint. Use the properties of an exponential model (see Sec. 3.5).

3.60*. Using Problems 3.47 and 3.58 and applying two one-sided critical regions, construct an unbiased test for verifying the hypothesis H_0: $\theta = \theta_0$ about the mean against the two-sided alternative H_1: $\theta \neq \theta_0$. Is this test uniformly most powerful?

3.61*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the normal distribution $\mathcal{N}(\mu, \theta^2)$. Construct a u.m.p. unbiased test for verifying a simple hypothesis H_0: $\theta = \theta_0$ against the two-sided alternative H_1: $\theta \neq \theta_0$.

Hint. Apply Theorem 4.5 [7, p. 196] on the general form of a u.m.p. unbiased test and use the solution to Problem 3.50.

3.62. Given a sample of size n from the distribution $\Gamma(\theta, 1)$, construct a u.m.p. unbiased test for verifying the hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta \neq \theta_0$.

Hint. Use the solution to Problems 3.43 and 3.61.

3.63. Given a large sample of size n, construct a local most powerful test for verifying the hypothesis H_0: $\theta = \theta_0$ against the common alternative H_1: $\theta \neq \theta_0$ for the model $Bi(k, \theta)$. Show that its power function $W_n(\theta)$ satisfies the limiting relation

$$\lim_{n\to\infty} W_n(\theta^{(n)}) = \Phi\left(\frac{-\beta\sqrt{k}}{\sqrt{\theta_0(1-\theta_0)}} + u_{\alpha/2}\right) + \Phi\left(\frac{\beta\sqrt{k}}{\sqrt{\theta_0(1-\theta_0)}} + u_{\alpha/2}\right)$$

for the significance level α and local alternatives of the form $\theta = \theta^{(n)} = \theta_0 + \beta/\sqrt{n}$.

Hint. Use the general form

$$\mathscr{X}_{1\alpha} = \{|U(\mathbf{x};\, \theta_0)| \geqslant -u_{\alpha/2}\sqrt{ni(\theta_0)}\}$$

for the asymptotic (at large n) two-sided test for regular models, where $U(\mathbf{x};\ \theta)$ is the contribution function of the sample $\mathbf{X} = (X_1, \ldots, X_n)$, and $i(\theta)$ is Fisher's information function [7, p. 199]. Use the solutions to Problems 3.39 and 3.40.

3.64. Given a large sample of size n, construct a local most powerful test for verifying the hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta \neq \theta_0$ for the model $\Pi(\theta)$. Show that its power function $W_n(\theta)$ satisfies the limiting relation

$$\lim_{n\to\infty} W_n(\theta^{(n)}) = \Phi\left(-\frac{\beta}{\sqrt{\theta_0}} + u_{\alpha/2}\right) + \Phi\left(\frac{\beta}{\sqrt{\theta_0}} + u_{\alpha/2}\right)$$

for the significance level α and local alternatives of the form $\theta = \theta^{(n)} = \theta_0 + \beta/\sqrt{n}$.

Hint. Use the hint to Problem 3.63 and the solution to Problem 3.41.

Tests of Hypotheses and Confidence Estimation

Problems 3.65-72 are based on the principle of correspondence between confidence estimation problems and tests of hypotheses (see Sec. 3.6).

3.65. Using the confidence intervals constructed in Problems 2.119-120 for the parameters θ_1 and θ_2 in the normal model $\mathcal{N}(\theta_1, \theta_2^2)$, construct the test for verifying the null hypotheses H_0 against the alternatives H_1 for the cases

(1) H_0: $\theta_1 = \theta_{10}$, H_1: $\theta_1 > \theta_{10}$;
(2) H_0: $\theta_1 = \theta_{10}$, H_1: $\theta_1 < \theta_{10}$;
(3) H_0: $\theta_1 = \theta_{10}$, H_1: $\theta_1 \neq \theta_{10}$;
(4) H_0: $\theta_2 = \theta_{20}$, H_1: $\theta_2 > \theta_{20}$;
(5) H_0: $\theta_2 = \theta_{20}$, H_1: $\theta_2 < \theta_{20}$;
(6) H_0: $\theta_2 = \theta_{20}$, H_1: $\theta_2 \neq \theta_{20}$.

3.66. Using the solution to Problems 2.122-123, show that at a significance level α the hypothesis on the equality of the means of two normal models with known variances can be verified by the test

$$\mathscr{X}_{1\alpha} = \{(\mathbf{x}, \mathbf{y}): |\bar{x} - \bar{y}| \geqslant u_{1-\alpha/2}\sqrt{\sigma_1^2/n + \sigma_2^2/m}\}.$$

If the variances are unknown, the test has the form

$$\mathscr{X}_{1\alpha} = \left\{(\mathbf{x}, \mathbf{y}): |\bar{x} - \bar{y}| \geqslant t_{1-\alpha/2, n+m-2}\sqrt{\frac{n+m}{nm(n+m-2)}\,(nS^2(\mathbf{x}) + mS^2(\mathbf{y}))}\right\}.$$

3.67. Using the solution to Problem 2.125, show that the test

$$\mathscr{S}_{1\alpha} = \left\{(\mathbf{x}, \mathbf{y}) \colon \frac{n(m-1)}{m(n-1)} \frac{S^2(\mathbf{x})}{S^2(\mathbf{y})} \leqslant F_{\alpha/2, n-1, m-1}\right\}$$

or

$$\mathscr{S}_{1\alpha} = \left\{(\mathbf{x}, \mathbf{y}) \colon \frac{n(m-1)}{m(n-1)} \frac{S^2(\mathbf{x})}{S^2(\mathbf{y})} \geqslant F_{1-\alpha/2, n-1, m-1}\right\}$$

can be used to verify the hypothesis that the variances of two normal models are equal.

3.68. Under the conditions of Problem 2.127, construct a test for verifying the homogeneity hypothesis H_0: $\tau = \theta_2/\theta_1 = 1$ (i.e., $\theta_1 = \theta_2$) and calculate its power function.

3.69. Using the confidence interval from Problem 2.128, construct a test for verifying the hypothesis H_0: $\theta = \theta_0$ for the respective model. Calculate its power function and make sure that it is unbiased.

3.70. Using the results of Problem 2.129, construct a test for verifying the hypothesis H_0: $\theta = \theta_0$ for the uniform distribution $R(0, \theta)$, calculate its power function, and make sure that it is unbiased.

3.71. Using the results of Problem 2.130, show that the test for verifying the hypothesis H_0: $\theta = \theta_0$ for Weibull's model $W(0, \lambda, \theta)$ has the form

$$\mathscr{S}_{1\alpha} = \left\{T \leqslant \frac{\theta_0^{\lambda}}{2} \chi^2_{\alpha_1, 2n}\right\} \cup \left\{T \geqslant \frac{\theta_0^{\lambda}}{2} \chi^2_{1-\alpha_2, 2n}\right\}, \quad \alpha_1 + \alpha_2 = \alpha.$$

In order to obtain an unbiased test, the quantities $\chi^2_{\alpha_1, 2n}$ and $\chi^2_{1-\alpha_2, 2n}$ are chosen as in Problem 3.62.

3.72. Using the condition of Problem 2.131 and its result, construct a test for verifying the hypothesis H_0: $(\theta_1, \theta_2) = (\theta_{10}, \theta_{20})$.

Likelihood Ratio Test

3.73*. Construct a likelihood ratio test for the hypothesis H_0: $\theta_1 = \theta_{10}$ for the mean of the normal model $\mathcal{N}(\theta_1, \theta_2^2)$ and show that for large samples it has the form

$$\mathscr{S}_{1\alpha} = \{\mathbf{x} \colon \sqrt{n-1}\,|\bar{x} - \theta_{10}|/S(\mathbf{x}) \geqslant -u_{\alpha/2}\},$$

and its power under the alternative $\theta_1^{(n)} = \theta_{10} + \beta/\sqrt{n}$ is equal to $1 - F_1(u^2_{\alpha/2};\, \beta^2/\theta_2^2)$ as $n \to \infty$ (see Sec. 3.7).

Hint. Use Problems 1.47 and 2.44, and the asymptotic theory for likelihood ratio tests [7, p. 212].

3.74*. Show that the likelihood ratio test for the hypothesis H_0: $\theta_2 = \theta_{20}$ for the variance of the normal model $\mathcal{N}(\theta_1, \theta_2^2)$ has the form

$$\mathscr{X}_{1\alpha} = \{nS^2(\mathbf{x})/\theta_{20}^2 \leqslant \chi^2_{\alpha_1, n-1}\} \cup \{nS^2(\mathbf{x})/\theta_{20}^2 \geqslant \chi^2_{1-\alpha_2, n-1}\},$$

where $\alpha_1 + \alpha_2 = \alpha$, and S^2 is the sample variance for a sample of size n. Compute the test's power function and find α_1 and α_2 for which the test is unbiased.

Hint. Use the fact that $\mathscr{L}_\theta(nS^2(\mathbf{X})/\theta_2^2) = \chi^2(n-1)$ and also the solution to Problem 3.61.

3.75. Construct a likelihood ratio test for the hypothesis H_0: $\theta = \theta_0$ for the model $Bi(1, \theta)$ and show that its asymptotic (for large samples) variant coincides with the local most powerful test constructed in Problem 3.63 (for $k = 1$).

Hint. Use the general theory of likelihood ratio tests for a polynomial distribution [7, pp. 207-208].

3.76. Construct the likelihood ratio test for the hypothesis H_0: $\theta = \theta_0$ for the model $\Pi(\theta)$ and make sure that its asymptotic variant for large samples coincides with the local most powerful test constructed in Problem 3.64.

Hint. Use the fact that as $n \to \infty$ the limiting distributions of the statistics $-2 \ln \lambda_n$ and $Q_n^{(2)} = U_n^2(\theta_0)/ni(\theta_0)$ coincide under the hypothesis H_0 [7, p. 207].

3.77*. Let $\overline{X_1}, \ldots, \overline{X_k}$ be the sample means of the independent samples with sizes $n_1, \ldots, n_k$ from the populations $Bi(1, \theta_1), \ldots, Bi(1, \theta_k)$, respectively. Construct and calculate an asymptotic (as $n_1, \ldots, n_k \to \infty$) likelihood ratio test for the homogeneity hypothesis H_0: $\theta_1 = \ldots = \theta_k$. Show that the test is similar to the χ^2-homogeneity test [7, pp. 160-161].

Hint. Use the solution to Problem 3.75.

3.78*. Let $\mathbf{X}_j = (X_{j1}, \ldots, X_{jn_j})$, $j = 1, \ldots, k$, be independent samples from the populations $\Pi(\theta_1), \ldots, \Pi(\theta_k)$, respectively. Construct and calculate an asymptotic (as $n_1, \ldots, n_k \to \infty$) likelihood ratio test for the homogeneity hypothesis H_0: $\theta_1 = \ldots = \theta_k$. Analyze the following data: the sums of the four samples of sizes 120, 100, 100, 125 from Poisson's populations were respectively 251, 323, 180, 426. Can we infer that the general means are equal?

3.79*. Let n_j, $\overline{X_j}$, and S_j^2 be the size, mean, and variance, respectively, of a sample from the population $\mathcal{N}(\theta_{1j}, \theta_2^2)$, $j = 1, \ldots, k$ (the samples are assumed to be independent). Construct a likelihood ratio test for the homogeneity hypothesis H_0: $\theta_{11} = \ldots = \theta_{1k}$. Show that in the

case of two samples ($k = 2$) the test has the form $\mathscr{X}_{1\alpha} = \{|T| \geqslant t_{1-\alpha/2, n_1+n_2-2}\}$ (compare with Problem 3.66), where

$$T = (\overline{X}_1 - \overline{X}_2)\sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2)(n_1 S_1^2 + n_2 S_2^2)}}.$$

Hint. Use Problems 2.86 and 2.114, and also the assertion $\mathscr{L}(T|H_0) = S(n - 2)$ [7, Theorem 1.12].

3.80*. Let $S_1^2, \ldots, S_k^2$ be sample variances constructed from independent samples with sizes $n_1, \ldots, n_k$ from the populations $\mathscr{N}(\theta_{1j}, \theta_{2j}^2)$, $j = 1, \ldots, k$, respectively. Construct a likelihood ratio test for the hypothesis H_0: $\theta_{21} = \ldots = \theta_{2k}$ that the variances are equal. Show that in the case of two samples ($k = 2$) the test has the form (compare with Problem 3.67)

$$\mathscr{X}_{1\alpha} = \{F \leqslant F_{\alpha_1, n_1-1, n_2-1}\} \cup \{F \geqslant F_{1-\alpha_2, n_1-1, n_2-1}\},$$

where $\alpha_1 + \alpha_2 = \alpha$, $F = [n_1(n_2 - 1)S_1^2]/[n_2(n_1 - 1)S_2^2]$.

Hint. Use the solution to Problem 3.79 and the assertion $\mathscr{L}(F|H_0) = S(n_1 - 1, n_2 - 1)$ [7, Theorem 1.13].

Various Problems

3.81. Assume that the observable random variables $X_1, \ldots, X_n$ are independent and normal though, generally speaking, they have various distributions. We test the hypothesis H_0 that they are distributed similarly. Using Problem 1.58, show that the critical region at the significance level α can be defined as $\mathscr{X}_{1\alpha} = \{|\eta| > v_\alpha\}$, where v_α is found from the beta distribution function by the relation $\mathrm{B}\left(1 - v_\alpha^2; \frac{n-2}{2}, \frac{1}{2}\right) = \alpha$.

We can also use published tables for the beta distribution $B\left(\frac{n-2}{2}, \frac{1}{2}\right)$.

3.82. Let $X_i = (X_{i1}, X_{i2})$, $i = 1, \ldots, n$, be independent observations on the two-dimensional random variable $\xi = (\xi_1, \xi_2)$ distributed normally with unknown parameters, and let ϱ_n be the sample correlation coefficient constructed from these data.

Using the results of Problem 1.59, show that the critical region

$$\mathscr{X}_{1\alpha} = \left\{|\varrho_n| \geqslant \frac{t_{1-\alpha/2, n-2}}{\sqrt{n - 2 + t_{1-\alpha/2, n-2}^2}}\right\}$$

defines a test of significance level α for the hypothesis H_0 that the components ξ_1 and ξ_2 are independent.

3.83. Let the observable random variables $X_1, \ldots, X_n$ be independent and $\mathscr{L}(X_i) = \Pi(\theta_i)$, $i = 1, \ldots, n$. Using the results obtained in Problem 1.60, show that for large n the test $\mathscr{X}_{1\alpha} = \{|T_n| \geqslant -u_{\alpha/2}\}$ can be applied to verify the homogeneity hypothesis H_0: $\theta_1 = \ldots = \theta_n$.

3.84. What kind of a goodness of fit test can be constructed from the result of Problem 1.61?

3.85*. (*Asymptotic efficiency of tests.*) Let us verify a simple hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta > \theta_0$ for a model with a scalar parameter $\theta \in \Theta$, where Θ is an interval on a real axis. We use a test of the form $\mathscr{X}_1 = \{T_n \geqslant \gamma_n\}$, where T_n is a statistic for a sample of size n, which possesses the following properties:

(a) there are functions $\mu(\theta)$ and $\sigma(\theta) > 0$, with

$$\mathscr{L}_\theta(T_n) \sim \mathscr{N}(\mu(\theta), \sigma^2(\theta)/n),$$

which are uniformly distributed in θ and lying in the interval $\theta_0 \leqslant \theta \leqslant \theta_0 + \eta$, where $\eta > 0$ is any number and $n \to \infty$;

(b) $\mu(\theta)$ is differentiable at the point θ_0 and $\mu'(\theta_0) > 0$, while $\sigma(\theta)$ is continuous at θ_0.

Prove that (1) given a significance level α, the critical boundary γ_n has the asymptotic form

$$\gamma_n = \mu(\theta_0) - u_\alpha \sigma(\theta_0)/\sqrt{n};$$

(2) for close alternatives of the form $\theta^{(n)} = \theta_0 + \beta/\sqrt{n}$, $\beta > 0$, the power $W_n(\theta^{(n)})$ of the test satisfies the limiting relation

$$e(\beta, \alpha) \equiv \lim_{n \to \infty} W_n(\theta^{(n)}) = \Phi(\beta\mu'(\theta_0)/\sigma(\theta_0) + u_\alpha).$$

Remark. The quantity $e = e(\beta, \alpha)$ is called *Pitman's efficiency* of the test $\mathscr{X}_{1\alpha} = \{\mathscr{T}_n \geqslant \mu(\theta_0) - u_\alpha\sigma(\theta_0)/\sqrt{n}\}$ and is frequently used as a measure for comparing various tests. For large samples the measure e describes the local behaviour of the test's power curve in the neighbourhood of the point θ_0.

3.86*. (Continued from Problem 3.85.) Let $T_n^{(j)}$, $j = 1, 2$, be two statistics meeting the conditions formulated above. We will label their characteristics with the superscript j. We assume that for each n there is an integer N_n such that

$$W_n^{(1)}(\theta_0 + \beta/\sqrt{n}) = W_{N_n}^{(2)}(\theta_0 + \beta/\sqrt{n}),$$

i.e., the powers of the tests are equal under the alternative $\theta^{(n)}$ if n

is the sample size in the first case, and N_n is the sample size in the second case. Suppose also that $N_n \to \infty$ at $n \to \infty$. Prove that

$$\lambda = \lim_{n\to\infty} \frac{n}{N_n} = \left(\frac{\mu_2'(\theta_0)}{\sigma_2(\theta_0)}\right)^2 \Bigg/ \left(\frac{\mu_1'(\theta_0)}{\sigma_1(\theta_0)}\right)^2 = \frac{e_2'}{e_1'},$$

where $e' = (\mu'(\theta_0)/\sigma(\theta_0))^2$.

Remark. The quantity e' is an increasing function of Pitman's efficiency $e(\beta, \alpha)$ for fixed β and α and can serve as a measure of the asymptotic efficiency of a test. This means that the relative efficiency of the second test with respect to the first one is equal to the limit of the ratio of the first sample size to the second one. The sample sizes are chosen so that for the indicated alternatives $\theta^{(n)}$ the test powers are equal.

3.87. We test the hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta > \theta_0$ for the normal model $\mathcal{N}(\theta, \sigma^2)$. Construct tests of the form $\mathscr{X}_1 = \{T_n \geqslant \gamma_n\}$ based on the statistics $T_n^{(1)} = \overline{X}$ (the sample mean) and $T_n^{(2)} = Z_{n,1/2}$ (the sample median) and show that the relative efficiency of the second test with respect to the first one is $\lambda = 2/\pi = 0.637\ldots$.

Hint. Use the solutions to Problems 3.86, 3.85, and 1.32.

3.88*. Suppose that we observe a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ distributed as $\mathscr{L}(\mathbf{X}) = \mathcal{N}(\theta\mathbf{t}, \Sigma = \|\sigma_{ij}\|_1^n)$, where θ is an unknown scalar parameter, $\mathbf{t} = (t_1, \ldots, t_n)$, $0 < t_1 < t_2 < \ldots < t_n$, are known constants, and $\sigma_{ij} = t_i$, $i \leqslant j$. (If $\eta(t)$, $t \geqslant 0$, is Wiener's process [2], i.e., a homogeneous random process with independent increments, and $\mathscr{L}(\eta(t)) = \mathcal{N}(\theta t, t)$, then $X_i = \eta(t_i)$, $i = 1, \ldots, n$, i.e., $\mathbf{X}$ are the observations on $\eta(t)$ at the moments $t_1, \ldots, t_n$.)

Show that the last observation X_n is a sufficient statistic for θ and, using this fact, construct the tests for verifying the hypothesis H_0: $\theta = 0$ that the process has no systematic trend (shift). Consider the alternatives H_1^+: $\theta > 0$, H_1^-: $\theta < 0$, and H_1: $\theta \neq 0$.

Hints. (1) Use the factorization test and establish the equation $\mathbf{t}\Sigma^{-1} = (0\ldots01)$.

(2) Use the solutions to Problems 3.47, 3.58, and 3.60.

CHAPTER 4

Linear Regression and the Least Squares Method

4.1. A *linear regression model* implies that the observable random variables $X_1, \ldots, X_n$ are "on average" linearly dependent on non-random factors $z_1, \ldots, z_k$, $k < n$, whose values may change from trial to trial. In this case the original statistical data are a set of the observed "responses" $X_1, \ldots, X_n$ and the respective factors, i.e., have the form $(x_i; z_1^{(i)}, \ldots, z_k^{(i)})$, $i = 1, \ldots, n$. We also assume that

$$\mathbf{E}X_i = \sum_{j=1}^{k} z_j^{(i)}\beta_j = \mathbf{z}^{(i)\prime}\boldsymbol{\beta}, \quad \mathbf{z}^{(i)} = (z_1^{(i)}, \ldots, z_k^{(i)}),$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ is the set of all unknown parameters called *regression coefficients.* If we introduce the random variables $\varepsilon_i = X_i - \mathbf{z}^{(i)}\boldsymbol{\beta}$, which are called the *measurement "errors"* and the *plan matrix* $\mathbf{Z} = \|\mathbf{z}^{(1)} \ldots \mathbf{z}^{(n)}\|$ sized $k \times n$, then we will obtain a matrix form

$$\mathbf{X} = \mathbf{Z}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbf{X} = (X_1, \ldots, X_n), \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n), \tag{4.1}$$

of the linear regression model. Here $\mathbf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and it is usually assumed that the random variables are uncorrelated and have the same variance, i.e., the matrix of the second moments of the observation vector $\mathbf{X}$ has the form

$$\mathbf{D}(\mathbf{X}) = \mathbf{D}(\boldsymbol{\varepsilon}) = \mathbf{E}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' = \sigma^2\mathbf{E}_n. \tag{4.2}$$

The quantity σ^2 is called the *residual variance* which is usually also unknown. If the non-random variables have the form $z_j = a_j(t)$, where $a_j(t)$ is a polynomial, we have *parabolic regression.*

The case of $k = 2$ is frequently applied. Here the vectors $\mathbf{z}^{(i)}$ are of the form $\mathbf{z}^{(i)} = (1, t_i)$, i.e., $\mathbf{E}X_i = \beta_1 + \beta_2 t_i$, $i = 1, \ldots, n$ (the average number of observations is a linear function of only the factor t). This is a *simple regression model,* where the straight line $\varphi(t) = \beta_1 + \beta_2 t$ is called a *regression line*, and the coefficient β_2 is its *slope.*

Some problems in regression analysis are solved under an additional assumption about the distribution of the errors ε and they are mostly taken to be normally distributed as $\mathscr{L}(\varepsilon) = \mathscr{N}(\mathbf{0}, \sigma^2\mathbf{E}_n)$. In this case the model has the form

$$\mathscr{L}(\mathbf{X}) = \mathscr{N}(\mathbf{Z}'\boldsymbol{\beta}, \sigma^2\mathbf{E}_n) \tag{4.3}$$

and is called the *normal regression*.

When seeking the vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ of unknown parameters, we use a linear regression model. We cannot measure the unknown parameters directly and only define some functions of them. Restoring a functional dependence belongs to this kind of problem. The expansion coefficients of the restored function, given a certain system of functions, are then the unknown parameters.

4.2. Regression analysis mostly deals with estimation of the unknown parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ and σ^2 of the model (4.1-2) or, in the case of the normal regression (4.3), with their confidence estimation and testing the hypotheses about the parameters.

A general technique for estimating unknown regression coefficients $\boldsymbol{\beta}$ is the *least squares method.* The estimates are found from the condition that the quadratic form

$$S(\boldsymbol{\beta}) = S(\mathbf{X}; \boldsymbol{\beta}) = (\mathbf{X} - \mathbf{Z}'\boldsymbol{\beta})'(\mathbf{X} - \mathbf{Z}'\boldsymbol{\beta}) \tag{4.4}$$

is minimized. The point $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_k)$ which satisfies the equation $S(\hat{\boldsymbol{\beta}}) = \min\limits_{\boldsymbol{\beta}} S(\boldsymbol{\beta})$ is called the *least squares estimate* (l.s.e.) for the parameter $\boldsymbol{\beta}$.

The matrix $\mathbf{A} = \mathbf{Z}\mathbf{Z}'$ is fundamental to these problems. We will assume that the matrix is non-singular (or, which is equivalent, rank $\mathbf{Z} = k$). Then the l.s.e. is uniquely defined by the *normal equation* $\mathbf{A}\boldsymbol{\beta} = \mathbf{Y} = \mathbf{Z}\mathbf{X}$ and has the form $\hat{\boldsymbol{\beta}} = \mathbf{A}^{-1}\mathbf{Y} = \mathbf{A}^{-1}\mathbf{Z}\mathbf{X}$. The estimate $\hat{\boldsymbol{\beta}}$ is unbiased ($\mathbf{E}\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$) and has a minimal variance (i.e., the variances of all the components of the vector $\hat{\boldsymbol{\beta}}$ are minimal) in the class of all linear (i.e., linearly dependent on the observations $\mathbf{X}$) unbiased estimates for $\boldsymbol{\beta}$. Moreover, any function $\hat{\mathbf{t}} = \mathbf{T}\hat{\boldsymbol{\beta}}$ possessing these properties is an estimate for the parameter $\mathbf{t} = \mathbf{T}\boldsymbol{\beta}$, where $\mathbf{T}$ is a given $m \times k$ matrix. Here $\mathbf{D}(\hat{\mathbf{t}}) = \sigma^2\mathbf{T}\mathbf{A}^{-1}\mathbf{T}'$ and, specifically, $\mathbf{D}(\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{A}^{-1}$.

The statistic

$$\bar{\sigma}^2 = \frac{1}{n-k} S(\hat{\boldsymbol{\beta}}) = \frac{1}{n-k}\mathbf{X}'\mathbf{B}\mathbf{X}, \quad \mathbf{B} = \mathbf{E}_n - \mathbf{Z}'\mathbf{A}^{-1}\mathbf{Z} \tag{4.5}$$

is an unbiased estimate for the residual variance σ^2 [7, pp. 223-226].

In the interpolation problems dealing with an unknown function

$x = f(t)$ which relates the variables t and x by the observations $(t_i, X_i = x_i + \varepsilon_i)$, $x_i = f(t_i)$, $i = 1, \ldots, n$, we seek an *interpolation polynomial* of the form

$$\varphi(t; \beta) = \sum_{j=1}^{k} \beta_j a_j(t),$$

where *Chebyshev's orthogonal polynomials* are used for $a_1(t)$, $a_2(t)$, ... The l.s.e.'s for the unknown coefficients β_i are calculated from the formulas

$$\hat{\beta}_j = \frac{1}{a_j^2} \sum_{i=1}^{n} a_j(t_i) X_i, \; a_j^2 = \sum_{i=1}^{n} a_j^2(t_i), \quad j = 1, 2, \ldots, \tag{4.6}$$

where the quantity $S(\hat{\beta}) = \sum_{i=1}^{n} X_i^2 - \sum_{j=1}^{k} a_j^2 \hat{\beta}_j^2$ defines the approximation accuracy. The first three Chebyshev's polynomials are of the form

$$a_1(t) \equiv 1, \quad a_2(t) = t - \bar{t}, \quad a_3(t) = (t - \bar{t})\left(t - \bar{t} - \frac{s_3}{s_2}\right) - s_2,$$

where

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i, \; s_k = s_k(t_1, \ldots, t_n) = \frac{1}{n} \sum_{i=1}^{n} (t_i - \bar{t})^k$$

(see [7, pp. 231-234]).

The least squares method is also applied when the dependence of $\mathbf{E}X_i$ on β is not linear. Suppose that

$$X_i = f(t_i, \beta_1, \ldots, \beta_k) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = \sigma^2$, $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$.

Then the l.s.e. $\hat{\beta}$ for the parameter β minimizes the expression

$$Q(\beta) = \sum_{i=1}^{n} (X_i - f(t_i, \beta_1, \ldots, \beta_k))^2$$

with respect to β.

Thus, $\hat{\beta}$ is a solution to the system

$$\frac{\partial Q(\beta)}{\partial \beta_i} = 0, \quad i = 1, \ldots, k.$$

We now calculate the estimates for β. Let us seek the unknown coefficients $(\beta_1, \beta_2, \beta_3)$ of the functional dependence

$$x(t) = \beta_1 + t\beta_2 + t^3\beta_3.$$

We will assume that the values of the function $x(t)$ have been found at the points $t_i = 2 + 3i/n$, $i = 1, \ldots, n$. We also assume that the measurement errors ε_i, $i = 1, \ldots, n$, are independent and normally distributed with $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = \sigma^2$. We then obtain a linear model

$$X_i = x(t_i) + \varepsilon_i = \beta_1 + t_i\beta_2 + t_i^3\beta_3 + \varepsilon_i, \; i = 1, \ldots, n,$$

and the estimates $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ satisfy the system of equations

$$\begin{cases} \sum_{i=1}^{n} (X_i - \hat{\beta}_1 - t_i\hat{\beta}_2 - t_i^3\hat{\beta}_3) = 0, \\ \sum_{i=1}^{n} (X_i - \hat{\beta}_1 - t_i\hat{\beta}_2 - t_i^3\hat{\beta}_3)\,t_i = 0, \\ \sum_{i=1}^{n} (X_i - \hat{\beta}_1 - t_i\hat{\beta}_2 - t_i^3\hat{\beta}_3)\,t_i^3 = 0. \end{cases} \tag{4.7}$$

Let $\beta_1 = 3$, $\beta_2 = -1$, $\beta_3 = 1$, $\sigma^2 = 0.04$. We simulate ε_i for $n = 25$ $n = 100$ and find the respective X_i. System (4.7) gives

(1) $n = 25$: $\hat{\beta}_1 = 2.983$, $\hat{\beta}_2 = -0.828$, $\hat{\beta}_3 = 0.895$, $\bar{\sigma}^2 = 0.034$.

(2) $n = 100$: $\hat{\beta}_1 = 2.992$, $\hat{\beta}_2 = -1.007$, $\hat{\beta}_3 = 1.001$, $\bar{\sigma}^2 = 0.046$.

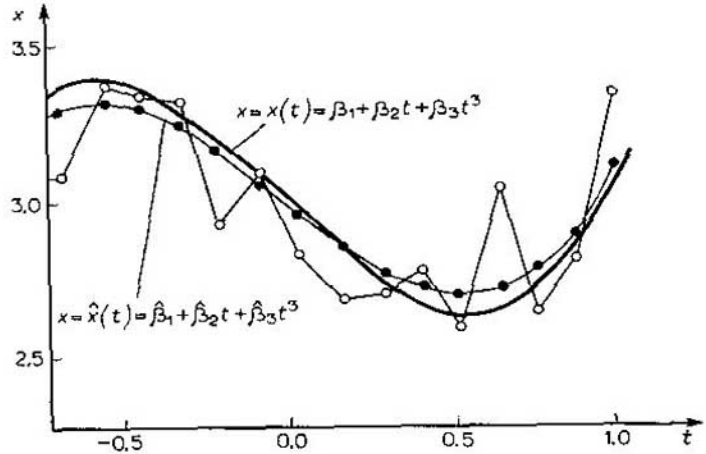


Fig. 5

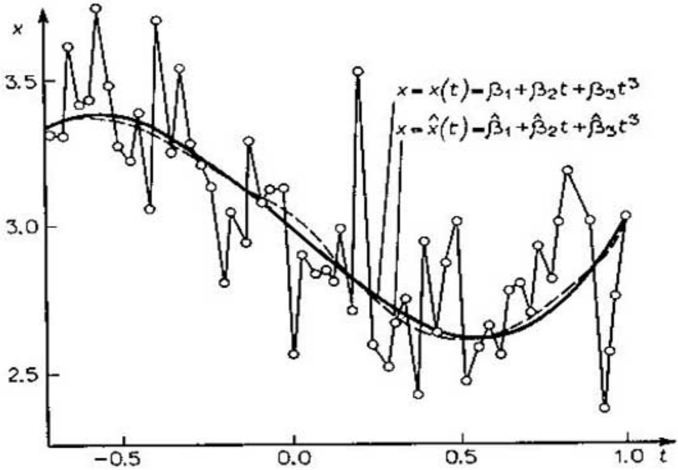


Fig. 6

Figures 5 and 6 give the exact curves of the function $x(t) = \beta_1 + \beta_2 t + \beta_3 t^3$ for $n = 25$ and $n = 100$, respectively. The sign ○ labels the measurement results for (t_i, X_i). The curves for $\hat{x}(t) = \hat{\beta}_1 + \hat{\beta}_2 t + \hat{\beta}_3 t^3$ are also shown.

4.3. In the normal regression scheme (4.3) the l.s.e.'s $\hat{\boldsymbol{\beta}}$ coincide with the maximum likelihood estimates (m.l.e.'s) for the parameters $\boldsymbol{\beta}$. The γ-confidence interval for the parameter β_j has the form

$$\left(\hat{\beta}_j \pm t_{(1+\gamma)/2,\, n-k} \sqrt{\frac{a^{jj}}{n-k} S(\hat{\boldsymbol{\beta}})}\right), \tag{4.8}$$

where a^{jj} is the jth diagonal element of the matrix $\mathbf{A}^{-1}$. For the residual variance we have

$$S(\hat{\beta})/\chi^2_{(1+\gamma)/2,\, n-k} < \sigma^2 < S(\hat{\beta})/\chi^2_{(1-\gamma)/2,\, n-k}. \tag{4.9}$$

The γ-confidence region for the vector $\mathbf{t} = \mathbf{T}\boldsymbol{\beta}$, where $\mathbf{T}$ is a given $m \times k$ matrix with rank $\mathbf{T} = m$, is constructed from the formula

$$\mathscr{G}_\gamma(\mathbf{X}) = \left\{\mathbf{t}\colon (\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{t})'\mathbf{D}^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{t}) < \frac{m}{n-k} S(\hat{\boldsymbol{\beta}}) F_{\gamma, m, n-k}\right\}, \tag{4.10}$$

where $\mathbf{D} = \mathbf{T}\mathbf{A}^{-1}\mathbf{T}'$ [7, pp. 237-238].

If we have to estimate simultaneously some linear combinations of the parameters $\boldsymbol{\beta}$, i.e., the quantities $\boldsymbol{\lambda}_r'\boldsymbol{\beta}$, $r = 1, \ldots, m$, where $\boldsymbol{\lambda}_r$ are given vectors, then the system of the joint confidence intervals with a confidence level equal to or greater than γ, has the form

$$\boldsymbol{\lambda}_r'\hat{\boldsymbol{\beta}} - u_\gamma(\mathbf{X};\ \boldsymbol{\lambda}_r) < \boldsymbol{\lambda}_r'\boldsymbol{\beta} < \boldsymbol{\lambda}_r'\hat{\boldsymbol{\beta}} + u_\gamma(\mathbf{X};\ \boldsymbol{\lambda}_r), \quad r = 1, \ldots, m, \tag{4.11}$$

where

$$u_\gamma(\mathbf{X};\ \boldsymbol{\lambda}) = \left[\frac{k}{n-k} S(\hat{\boldsymbol{\beta}}) F_{\gamma, k, n-k}(\boldsymbol{\lambda}'\mathbf{A}^{-1}\boldsymbol{\lambda})\right]^{1/2}$$

(see [7, p. 241]).

Finally, to test a *linear hypothesis* of the form H_0: $\boldsymbol{\beta} \in \mathbf{B}_0 = \{\boldsymbol{\beta}: \mathbf{T}\boldsymbol{\beta} = \mathbf{t}_0\}$, where $\mathbf{T}$ is a given $m \times k$ matrix with rank $\mathbf{T} = m$, and $\mathbf{t}_0$ is a given vector, we use an *F-test* with a critical region of the form

$$\mathscr{X}_{1\alpha} = \left\{\frac{n-k}{m} \frac{S_\mathbf{T} - S(\hat{\boldsymbol{\beta}})}{S(\hat{\boldsymbol{\beta}})} \geqslant F_{1-\alpha, m, n-k}\right\}, \tag{4.12}$$

where $S_\mathbf{T} = \min\limits_{\boldsymbol{\beta}:\mathbf{T}\boldsymbol{\beta} = \mathbf{t}_0} S(\boldsymbol{\beta})$ is a conditional (under the hypothesis H_0) minimum of $S(\boldsymbol{\beta})$ [7, pp. 242-243].

Problems

4.1. Given a linear model (4.1) for $k = 2$, write an explicit expression for the l.s.e. $(\hat{\beta}_1, \hat{\beta}_2)$ through $(X_1, \ldots, X_n)$ and $\mathbf{z}^{(i)} = (z_1^{(i)}, z_2^{(i)})$, $i = 1, \ldots, n$.

4.2. Given a simple regression model

$$X_i = \beta_1 + \beta_2 t_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

find an explicit form for $(\hat{\beta}_1, \hat{\beta}_2)$, check whether they are unbiased, and find the condition for them to be consistent.

4.3. Calculate the estimate $\tilde{\sigma}^2$ (see (4.5)) for the residual variance σ^2 in Problem 4.2. Find the sufficient condition for it to be consistent.

4.4. Find cov $(\hat{\beta}_1, \hat{\beta}_2)$ for the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ from Problem 4.2.

4.5. The values of the function $x(t) = \beta_1 + \beta_2 t + \beta_3 t^2$ have been measured at the points t_i, $i = 1, \ldots, n$, i.e.,

$$X_i = \beta_1 + \beta_2 t_i + \beta_3 t_i^2 + \varepsilon_i, \quad \mathbf{E}\varepsilon_i = 0, \quad \mathbf{D}\varepsilon_i = \sigma^2.$$

Find (1) the l.s.e.'s $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ for the parameters β_1, β_2, β_3; (2) $\mathbf{E}\hat{\beta}_i$, $\mathbf{D}\hat{\beta}_i$, $i = 1, 2, 3$; cov $(\hat{\beta}_i, \hat{\beta}_j)$.

4.6. Is the statistic

$$\hat{I} = \int_a^b \hat{x}(t)\,dt$$

an unbiased estimate for the integral $I = \int_a^b x(t)\,dt$, where $\hat{x}(t) = \hat{\beta}_1 + \hat{\beta}_2 t + \hat{\beta}_3 t^2$, and $x(t)$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ are defined as in Problem 4.5? Find $\mathbf{D}\hat{I}$.

4.7. Simulate the observations $X_i = \beta_1 + \beta_2 t_i + \varepsilon_i$, $i = 1, \ldots, n$, if $n = 100$, $t_i = 2i/n$, $\beta_1 = 2$, $\beta_2 = 1$, and ε_i are independent random variables uniformly distributed on the segment $[-1.386, 1.386]$. Plot the functions $x(t) = 2 + t$ and $\hat{x}(t) = \hat{\beta}_1 + \hat{\beta}_2 t$ on the segment $[0, 2]$. Mark the points (t_i, X_i), $i = 1, \ldots, n$.

4.8. Solve Problem 4.7 for normally distributed ε_i with $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = 0.16$.

4.9. Construct a γ-confidence interval as in Problem 4.8 for the parameters β_1, β_2, and σ^2 (see (4.8), (4.9)) and a γ-confidence ellipse $\mathscr{G}_\gamma(\mathscr{X})$ (see (4.10)) for the vector $\boldsymbol{\beta} = (\beta_1, \beta_2)$. use $\gamma = 0.9$ and $\gamma = 0.95$.

Hint. Use the solutions to Problems 4.2 and 4.3.

4.10. Using independent measurements at the same points (X_i, t_i), $i = 1, \ldots, n$, of a linear function $x(t) = \beta_1 + \beta_2 t$ (measurement errors are normally distributed as $\mathcal{N}(0, \sigma^2)$ with an unknown variance), construct a confidence interval for the integral of this function on the segment $-a \leqslant t \leqslant a$ (a is given). Carry out the calculations for the data $(2.96, -2)$, $(3.20, -1)$, $(3.41, 0)$, $(3.63, 1)$, $(3.79, 2)$ for $a = 2$ and a confidence level $\gamma = 0.95$.

4.11. A point is moving uniformly in a straight line. The values of the coordinate $a(t)$ at the moments $t = 1, 2, 3, 4, 5$ are 12.98, 13.05, 13.32, 14.22, 13.97, respectively. Assuming that the measurement errors are independent and normally distributed as $\mathcal{N}(0, \sigma^2)$, construct a 95% confidence ellipse for the point $(a(0), v)$, where v is the speed of the point.

4.12. Simulate the observations $X_i = \beta_1 + \beta_2 t_i + \beta_3 t_i^2 + \varepsilon_i$, $i = 1, \ldots, n$, with $\beta_1 = -8$, $\beta_2 = 10$, $\beta_3 = -2$, $n = 100$, $t_i = 1 + 2i/n$, where ε_i are independent random variables uniformly distributed on the segment $[-1.386, 1.386]$. Plot the functions $x(t) = \beta_1 + \beta_2 t + \beta_3 t^2$ and $\hat{x}(t) = \hat{\beta}_1 + \hat{\beta}_2 t + \hat{\beta}_3 t^2$ on the segment $[1, 3]$. Mark the points (t_i, X_i), $i = 1, \ldots, n$.

4.13. Solve Problem 4.12 for normally distributed ε_i with $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = 0.16$.

4.14. In the previous problem construct γ-confidence intervals for the parameters β_1, β_2, β_3, and σ^2 (see (4.8-9)) and a system of joint confidence intervals for the level equal to or greater than γ for β_1, β_2, β_3 (see (4.11)).

Hint. Use the solution to Problem 4.5.

4.15. In a quadrangle $ABCD$ the results of independent measurements at the same points of the angles ABD, DBC, ABC, BCD, CDB, BDA, CDA, DAB (in degrees) are 50.78, 30.25, 78.29, 99.57, 50.42, 40.59, 88.87, 89.86. Assuming that the measurement errors are normally distributed as $\mathcal{N}(0, \sigma^2)$, find the l.s.e.'s for the angles $\beta_1 = ABD$, $\beta_2 = DBC$, $\beta_3 = CDB$, $\beta_4 = BDA$. Construct a 95% confidence interval for σ^2.

4.16*. Prove that the l.s.e. $\hat{\beta}$ is an optimum estimate for β in the class of all linear (i.e., linearly dependent on $\mathbf{X}$) unbiased estimates for β (i.e., the variances $\mathbf{D}\hat{\beta}_i$ are minimal for all i). Show that $\mathbf{D}(\hat{\beta}) = \sigma^2\mathbf{A}^{-1} = \sigma^2|a^{ij}|$ and $\sum_{i=1}^{k} \mathbf{D}\hat{\beta}_i = \sigma^2 \operatorname{tr} \mathbf{A}^{-1} = \sigma^2 \sum_{i=1}^{k} \lambda_i^{-1}$, λ_i are the eigenvalues of the matrix $\mathbf{A}$. Derive the consistency condition of the estimate $\hat{\beta}_i$: $\min \lambda_i \to \infty$ as $n \to \infty$.

4.17. Prove that $\tilde{\sigma}^2$ is an unbiased estimate for the residual variance σ^2. Obtain an explicit form for the dependence of $\tilde{\sigma}^2$ on $\mathbf{X}$ from formula (4.5). Obtain the formulas: $\mathbf{E}(\mathbf{U}) = \mathbf{0}$, $\mathbf{D}(\mathbf{U}) = \sigma^2\mathbf{B}$, $\operatorname{cov}(\mathbf{U}, \hat{\beta}) = \mathbf{0}$, where $\mathbf{U} = \mathbf{X} - \mathbf{Z}'\hat{\beta} = \mathbf{B}\mathbf{X}$.

Hint. Use the expansion $S(\beta) = S(\hat{\beta}) + (\hat{\beta} - \beta)'\mathbf{A}(\hat{\beta} - \beta)$, the formulas $\operatorname{cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 a^{ij}$ (Problem 4.16) and $\hat{\beta} = \beta + \mathbf{A}^{-1}\mathbf{Z}\varepsilon$, $\mathbf{U} = \mathbf{B}\varepsilon$.

4.18. Suppose that the plan matrix $\mathbf{Z}$ has orthogonal rows. Find the l.s.e.'s $\hat{\beta}_1, \ldots, \hat{\beta}_k$ and their second moments.

4.19*. We have k items with unknown weights $\beta_1, \ldots, \beta_k$. In order to find the weights, we weigh items in combinations. Each operation consists of putting a few items on one pan and a few items on the other, which is then balanced with an additional weight. We obtain the relations

$$z_1^{(i)}\beta_1 + \ldots + z_k^{(i)}\beta_k = y_i$$

(for the ith weighing, $i = 1, \ldots, n$), where $z_j^{(i)} = 1, -1, 0$ (depending on whether the jth item is on the left pan, on the right pan, or not weighed), and y_i is the additional weight. Assuming that the measurement errors are independent and normally distributed as $\mathcal{N}(0, \sigma^2)$, estimate the weights of four items using the following table for eight weighings:

β_1	1	1	1	1	1	1	1	1
β_2	1	−1	1	−1	1	−1	1	−1
β_3	1	1	−1	−1	1	1	−1	−1
β_4	1	−1	−1	1	1	−1	−1	1
Weight	20.2	8.1	9.7	1.9	19.9	8.3	10.2	1.8

Find the covariance matrix for the estimates, and the estimate for σ^2. Compare the precision of these estimates with that of the estimates obtained by weighing every item a few times and finding the arithmetic mean of these values.

Hint. Use the solution to the previous problem.

4.20. For the data of Problem 4.19 construct a system of joint confidence intervals for $\beta_1, \ldots, \beta_4$ with a significance level $\geqslant 0.95$.

4.21. Find the maximum likelihood estimates for the parameters $\boldsymbol{\beta}$ and σ^2 of the normal regression in (4.3) and calculate their biases.

4.22. Show that the γ-confidence interval for an arbitrary linear combination $\boldsymbol{\lambda}'\boldsymbol{\beta} = \sum_{j=1}^{k} \lambda_j\beta_j$ of the normal regression coefficients in (4.3) has the form

$$\left(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} \pm t_{(1+\gamma)/2,\,n-k}\sqrt{\frac{1}{n-k}\,S(\hat{\boldsymbol{\beta}})(\boldsymbol{\lambda}'\mathbf{A}^{-1}\boldsymbol{\lambda})}\right).$$

4.23. Construct a γ-confidence interval for the ordinate $\varphi(t) = \beta_1 + \beta_2 t$ of a regression line at an arbitrary point t (the model is assumed to be normal). Make calculations for the data of Problem 4.8 for $t = 1.5$ and $\gamma = 0.95$.

Hint. Use the solutions to Problems 4.2-4 and 4.22.

4.24. Verify that the intervals

$$\left(\hat{\beta}_i - \hat{\beta}_j \pm \left[\frac{k}{n-k}\,S(\hat{\boldsymbol{\beta}})F_{\gamma,k,n-k}(a^{ii} - 2a^{ij} + a^{jj})\right]^{1/2}\right),$$
$$1 \leqslant j < i \leqslant k,$$

constitute a system of joint confidence intervals at a level greater than or equal to γ for the differences $\beta_i - \beta_j$, $i > j$.

4.25. Construct a system of joint confidence intervals for the mean values of all the observations $X_1, \ldots, X_n$ in the normal regression model.

4.26*. Let $\mathbf{T}$ be a given $m \times k$ matrix ($m \leqslant k$) with rank $\mathbf{T} = m$, and let $\mathbf{t}_0$ be a given m-dimensional vector such that the system $\mathbf{T}\boldsymbol{\beta} = \mathbf{t}_0$ is compatible. We write $S_{\mathbf{T}} = \min\limits_{\hat{\boldsymbol{\beta}}_{\mathbf{T}}:\ \mathbf{T}\boldsymbol{\beta} = \mathbf{t}_0} S(\hat{\boldsymbol{\beta}})$ and call the value of $\boldsymbol{\beta}$ for which $S_{\mathbf{T}} = S(\hat{\boldsymbol{\beta}}_{\mathbf{T}})$ the *generalized* l.s.e. $\hat{\boldsymbol{\beta}}_{\mathbf{T}}$. Prove that

$$\hat{\boldsymbol{\beta}}_{\mathbf{T}} = \hat{\boldsymbol{\beta}} - \mathbf{A}^{-1}\mathbf{T}'\mathbf{D}^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{t}_0),$$

where the matrix $\mathbf{D} = \mathbf{T}\mathbf{A}^{-1}\mathbf{T}'$ is positive definite. Find the expansion

$$S_{\mathbf{T}} = S(\hat{\boldsymbol{\beta}}) + (\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{t}_0)'\mathbf{D}^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}} - \mathbf{t}_0).$$

4.27. Show that the test of a significance level α for verifying the hypothesis H_0: $\beta_2 = \beta_{20}$ which fixes the slope of the regression line (in a normal model) is defined by the critical region

$$\mathscr{X}_{1\alpha} = \left\{ |\hat{\beta}_2 - \hat{\beta}_{20}| \geqslant t_{1-\alpha/2, n-2} \sqrt{S(\hat{\beta}) \Big/ \left[(n-2) \sum_{i=1}^{n} (t_i - \bar{t})^2 \right]} \right\}.$$

|*Hint.* Use the solutions to Problems 4.2-3 and relation (4.12).

4.28. Given the data of Problem 4.8, find the values of the significance level α for which the hypothesis H_0: $\beta_2 = 1.2$ should be rejected.

4.29. The values of independent random variables $X_j^{(i)}$, $i = 1, 2, 3, 4$; $j = 1, 2$, are given in the following table:

j \ i	1	2	3	4
1	8.67	9.71	10.16	13.65
2	10.03	10.23	9.26	13.79

Assuming that $\mathscr{L}(X_j^{(i)}) = \mathscr{N}(\mu_i, \sigma^2)$ (all the parameters are unknown), construct the estimates for μ_1, μ_2, μ_3, μ_4, and σ^2 and test the homogeneity hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (take the significance level 0.1).

4.30. Construct an interpolation polynomial of the form $\varphi_k(t, \beta) = \sum_{j=1}^{k} \beta_j a_j(t)$ for $k = 2$ and 3, where $a_j(t)$ are Chebyshev's

polynomials (see (4.6)), using the data for the unknown function $x = f(t)$ from the following table:

t_i	0.40	0.52	0.61	0.70	0.79	0.86	0.89	0.95	0.99
x_i	0.39	0.50	0.57	0.65	0.71	0.76	0.78	0.81	0.84

How does the precision of the interpolation change when we go from $k = 2$ to $k = 3$?

4.31. Prove a similar problem for the data

t_i	0	4	10	15	21	29	36	51	68
x_i	66.7	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

4.32. Compute the fourth polynomial in a system of Chebyshev's orthogonal polynomials.

Hint. Use the recurrence relation

$$a_{\tau+1}(t) = (t + \alpha)a_\tau(t) + \beta a_{\tau-1}(t),$$

where

$$\alpha = -\sum_{i=1}^{n} t_i a_\tau^2(t_i)/a_\tau^2, \quad \beta = -\sum_{i=1}^{n} t_i a_{\tau-1}(t_i)a_\tau(t_i)/a_{\tau-1}^2$$

(see (4.6)).

4.33. Simulate the observations $X_i = t_i^2 + \varepsilon_i$, $i = 1, \ldots, n$, if $n = 100$, $t_i = 2 + 0.1(i - 1)$, and ε_i are independent random variables uniformly distributed on the segment $[0, 0.7]$.

(1) Construct an interpolation polynomial $\varphi_k(t; \hat{\beta}) = \sum_{j=1}^{k} \hat{\beta}_j a_j(t)$ for $k = 2, 3, 4$, where $a_j(t)$, $j = 1, \ldots, 4$, are Chebyshev's polynomials.

(2) Plot the functions $x = t^2$, $x = \varphi_k(t; \hat{\beta})$, $k = 2, 3, 4$.

(3) How does the precision of the interpolation change with the growth of k?

4.34. Solve Problem 4.33 for normally distributed ε_i with $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = 0.04$.

4.35. Solve Problem 4.33 with $X_i = e^{t_i} + \varepsilon_i$.

4.36. Solve Problem 4.35 for normally distributed ε_i with $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = 0.04$.

4.37. Let $Y_1, \ldots, Y_n$ be independent random variables with a common distribution function $F_0((x - \beta_1)/\beta_2)$, where $F_0(x)$ is a known continuous distribution function, and the shift parameter β_1 and the scale parameter $\beta_2 > 0$ are unknown. Then $Y_j = \beta_1 + \beta_2 U_j$, where the random variables $U_1, \ldots, U_n$ are independent and their distribution function is $F_0(x)$. We write $Y_{(j)} = \beta_1 + \alpha_j \beta_2 + \beta_2 \varepsilon_j$ for the respective order statistics, where $\varepsilon_j = U_{(j)} - \alpha_j$, $\alpha_j = \mathbf{E}U_{(j)}$, $j = 1, \ldots, n$. Find the estimates for the parameters β_1, β_2 using the least squares method.

Hint. Here the random variables $\mathbf{Y} = (Y_{(1)}, \ldots, Y_{(n)})$ satisfy a model of linear regression with correlated observations, i.e., $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \operatorname{cov}(U_{(i)}, U_{(j)}) \equiv g_{ij}$ are known, and we may go to the uncorrelated variables $\mathbf{X} = \mathbf{G}^{-1/2}\mathbf{Y}$, where the matrix $\mathbf{G} = \|g_{ij}\|_1^n$ is assumed to be non-singular.

CHAPTER 5

Decision Functions

5.1. Suppose that we are given a sample space $\mathscr{X} = \{x\}$ of the values of the observable random variable X and a function $\delta(x)$ on it whose values are in the set $D = \{d\}$ of possible decisions made from an observation on some value of X. In this case $\delta(x)$ is called a *decision function (rule, procedure)*. Suppose also that $\mathscr{L}(X) \in \mathscr{F} = \{F(x; \theta), \theta \in \Theta\}$ and for every pair $(\theta, d) \in \Theta \times D$ a number $L(\theta, d) \geqslant 0$ is defined. This number is the loss due to making the decision d when X is distributed as $F(x; \theta)$. Then $L(\theta, d)$ is said to be a *loss function*. In point estimation, for example, decisions are estimates for the parameter θ, and therefore the decision set D usually coincides with the parametric set Θ, the decision function δ is called the estimate, and the loss $L(\theta, d)$ is the difference between the value of θ and the estimate of d. As a rule, such problems imply that $L(\theta, d) = \omega(|\theta - d|)$, where ω is a strictly monotone error function $|d - \theta|$.

The quantity $R(\theta, \delta) = \mathbf{E}_\theta L(\theta, \delta(X))$ is called a *risk function* of the procedure δ and characterizes the average loss due to the application of the decision rule δ when the observable random variable X is distributed as $F(x; \theta)$. If the condition

$$R(\theta, \delta') \leqslant R(\theta, \delta) \quad \forall \theta \in \Theta \tag{5.1}$$

is met for the two rules δ' and δ, the strict inequality holding for at least one θ, then the rule δ' is preferable. The rule δ is then called *inadmissible*. A decision rule which is not inadmissible (there is no preferable rule) is called *admissible*. In practice we restrict ourselves to the class of admissible decision rules no two of which can be compared in the sense of (5.1). In order to choose the best decision rule among the admissible ones, we use the Bayes and minimax approaches.

5.2. The *Bayes approach* implies that the parameter θ is a random variable with some (*a priori*) *distribution* $\mathscr{L}(\theta)$ given by the distribu-

tion density $\pi(\theta)$ (or probability in the discrete case). We may calculate the total average loss due to using the rule δ, i.e.,

$$r(\delta) = \int R(\theta, \delta)\pi(\theta)\, d\theta$$

$\left(\text{in the discrete case we have } \sum_i R(\theta_i, \delta)\pi(\theta_i)\right)$, which is called the *Bayes risk*. All decision rules can be ordered with respect to this value. The procedure δ^* which minimizes the Bayes risk $r(\delta)$ is an optimal rule called the *Bayes solution*.

We now suggest an algorithm to find the Bayes solution for a given a priori distribution of the parameter $\pi(\theta)$ [7, pp. 270-271]:

(a) We seek an *a posteriori distribution* $\pi(\theta|x)$ for $X = x$ using the formula $\pi(\theta|x) = f(x;\ \theta)\pi(\theta)/f(x)$, where $f(x) = \mathbf{E}f(x;\ \theta) = \int f(x;\ \theta)\pi(\theta)\, d\theta$ $\left(\text{or } \sum_i f(x;\ \theta_i)\pi(\theta_i) \text{ in the discrete case}\right)$.

(b) We calculate the average loss for the solution d with respect to the a posteriori distribution, viz.,

$$\mathbf{E}(L(\theta, d)|x) = \int L(\theta, d)\pi(\theta|x)\, d\theta \quad \left(\text{or } \sum_i L(\theta_i, d)\pi(\theta_i|x)\right).$$

(c) We choose the solution $d^* = \delta^*(x)$ for which the average loss is minimal.

5.3. When the a priori information about θ is absent, we use the *maximal risk* $m(\delta) = \sup_{\theta\in\Theta} R(\theta, \delta)$ to compare the admissible decision rules. The rule $\bar{\delta}$ minimizing $m(\delta)$ is considered to be the best one and is called a *minimax decision rule*. In some cases this rule can be constructed if there is an a priori distribution of the parameter $\pi(\theta) > 0$ for which the risk function of the respective Bayes rule δ^* is constant, i.e., $R(\theta, \delta^*) \equiv \text{const}$ (this distribution of π is called the *least favourable a priori distribution*) and then $\bar{\delta} = \delta^*$ [7, p. 271].

5.4. We now turn to an important special case when $\Theta = \{\theta_1, \ldots, \theta_k\}$, i.e., when only a finite number k of the distributions $F_i(x) = F(x;\ \theta_i)$, $i = 1, \ldots, k$, are possible for the observable random variable X, and, given an observation on X, we must choose one of them as the true distribution.

Problems of this kind are called *classification problems*. Here the set of possible decisions is $D = \{d_1, \ldots, d_k\}$, where d_i implies that the distribution F_i, $i = 1, \ldots, k$, should be chosen as the true distribu-

tion, and every decision rule $\delta(x)$ generates a partition of the sample space $\mathscr{X} = W_1 \cup \ldots \cup W_k$, $W_i \cap W_j = \varnothing$, $i \neq j$, where $W_i = \{x: \delta(x) = d_i\}$, $i = 1, \ldots, k$. The Bayes solution δ^* is defined by the partition $\mathscr{X} = W_1^* \cup \ldots \cup W_k^*$, where $W_i^* = \{x: h_i(x) = \min_{1 \leqslant j \leqslant k} h_j(x)\}$, $i = 1, \ldots, k$, and

$$h_j(x) = \sum_{i=1}^{k} l(j|i)\pi_i f_i(x),\ l(j|i) = L(\theta_i, d_j),\ \pi_i = \pi(\theta_i)$$

(if the minimum is attained for a few values of j, then we choose the smallest of them for the subscript i). If the loss is $l(j|i) = 1$, $j \neq i$, or is unknown, or cannot be estimated by a number, then the Bayes rule is replaced by the *principle of maximum a posteriori probability*, which requires that the *object with the observation x is placed in the class whose a posteriori probability* $\pi_i(x) = f_i(x)\pi_i \Big/ \sum_{s=1}^{k} \pi_s f_s(x)$, $i = 1, \ldots, k$, *is maximal*. In such cases [7, p. 276] we have

$$W_i^* = \{x: \pi_i f_i(x) = \max_{1 \leqslant j \leqslant k} \pi_j f_j(x)\},\ i = 1, \ \ldots, \ k.$$

In order to construct a minimax solution $\tilde{\delta}$, we seek the least favourable a priori distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$ from the condition that the components of the risk vector $\mathbf{R}(\delta^*) = (R_1(\delta^*), \ldots, R_k(\delta^*))$ of the respective Bayes solution are equal, where

$$R_i(\delta^*) = R(\theta_i, \delta^*) = \sum_{j=1}^{k} l(j|i)\mathbf{P}_{\theta_i}(X \in W_j^*),\quad i = 1, \ \ldots, \ k.$$

Problems

5.1. Let $\mathscr{L}(X) = Bi(1, \theta)$, $\Theta = \{\theta_1 = 1/3, \theta_2 = 2/3\}$, the decision set be $D = \{d_1, d_2\}$, and let the loss function $L(\theta_i, d_j)$ be given by the table

	d_1	d_2
θ_1	0	1
θ_2	2	0

(1) Find the admissible decision rules and the minimax solution among them.

(2) Find a Bayes solution δ^* for an arbitrary a priori distribution of $\pi(\theta_1) = \alpha$, $\pi(\theta_2) = 1 - \alpha$, $\alpha \in [0, 1]$, and plot the Bayes risk $\varrho(\alpha) = r(\delta^*)$ as a function of α.

5.2. Find all Bayes solutions for $\mathscr{L}(X) = Bi(1, \theta)$, $\Theta = \{\theta_1 = 1/4, \theta_2 = 3/4\}$, $D = \{d_1, d_2, d_3\}$, and let the loss function $L(\theta_i, d_j)$ be given by the table

	d_1	d_2	d_3
θ_1	0	1	1/2
θ_2	4	0	1/2

Plot the Bayes risk $\varrho(\alpha) = r(\delta^*)$ as a function of $\alpha = \pi(\theta_1)$, $\alpha \in [0, 1]$.

Hint. Compare the average loss with respect to the a posteriori distribution given in the solution to Problem 5.1 (2).

5.3. Let $\mathscr{L}(X) = Bi(1, \theta)$, $\Theta = \{\theta_1, \theta_2\}$, $D = \{d_1, d_2\}$, and let the loss function $L(\theta_i, d_j)$ be given by the table

	d_1	d_2
θ_1	0	a
θ_2	b	0

for $a, b > 0$. Consider the cases of $\theta_1 = 2/3$, $\theta_2 = 1/2$ and $\theta_1 = 3/4$, $\theta_2 = 1/2$.

Show that in both cases the sets of admissible decision rules coincide, but in the second case the Bayes solution is preferable for any a priori distribution of the parameter.

Hint. Use the solution to Problem 5.1.

5.4. Suppose that $\mathscr{L}(X) = Bi(3, \theta)$, $\Theta = \{\theta_1 = 10^{-2}, \theta_2 = 10^{-1}\}$, the decision set is $D = \{d_1, d_2\}$, and the loss function $L(\theta_i, d_j)$ is given by the table

	d_1	d_2
θ_1	0	2
θ_2	1	0

Consider the decision rules $\delta_i = (\delta_i(0), \delta_i(1), \delta_i(2), \delta_i(3))$, where

$\delta_1 = (d_1, d_2, d_2, d_2)$, $\delta_2 = (d_1, d_1, d_2, d_2)$, $\delta_3 = (d_1, d_1, d_1, d_2)$.

Show that the rules cannot be compared and find the minimax solution among them.

5.5. Suppose that $\mathscr{L}(X) = \overline{Bi}(1, \theta)$, $\Theta = \{\theta_1, \theta_2\}$, $D = \{d_1, d_2\}$, and the loss function is given by the table

	d_1	d_2
θ_1	0	a
θ_2	b	0

Find the minimax decision function among the functions

$$\delta_i(x) = \begin{cases} d_1 & \text{for} \quad x = 0, 1, \ldots, i-1, \\ d_2 & \text{for} \quad x = i, i+1, \ldots, \end{cases} \qquad i = 1, 2, \ldots.$$

5.6. Show that if in the previous problem we take the Poisson law $\Pi(\theta)$ instead of $\mathscr{L}(X)$, then for $a(1 - e^{-\theta_1}) \leqslant be^{-\theta_2}$ we will have $\bar{\delta} = \delta_1$, and the respective risk vector will be $(a(1 - e^{-\theta_1}), be^{-\theta_2})$.

5.7. Let X be a random variable distributed either as $F_1(x) = F(x; \theta_1)$ or as $F_2(x) = F(x; \theta_2)$. Let the decision set be $D = \{d_1, d_2\}$, and let the loss function be given by the table

	d_1	d_2
θ_1	0	a
θ_2	b	0

Construct a Bayes solution for a given a priori distribution of (π_1, π_2) and calculate the respective risk (compare with Problem 3.51). Consider the case when F_i is the normal distribution $\mathcal{N}(\theta_i, \sigma^2)$, $i = 1, 2$.

5.8*. Let the observable random variable X be distributed normally with an unknown mean θ and known variance σ^2, the decision set be $D = \{d_1, d_2, d_3\}$, and let the loss function $L(\theta, d)$ be given by the table

$\theta \setminus d$	d_1	d_2	d_3
< 0	0	1	2
$= 0$	1	0	1
> 0	2	1	0

Consider the decision functions

$$\delta_{ab}(x) = \begin{cases} d_1 & \text{for} \quad x < a, \\ d_2 & \text{for} \quad a \leqslant x \leqslant b, \\ d_3 & \text{for} \quad x > b, \end{cases}$$

where $a < 0 < b$.

Show that the risk function has the form

$$R(\theta, \delta_{ab}) = \begin{cases} \Phi\left(\dfrac{\theta - a}{\sigma}\right) + \Phi\left(\dfrac{\theta - b}{\sigma}\right) & \text{for} \quad \theta < 0, \\ \Phi(a/\sigma) + \Phi(-b/\sigma) & \text{for} \quad \theta = 0, \\ \Phi\left(\dfrac{a - \theta}{\sigma}\right) + \Phi\left(\dfrac{b - \theta}{\sigma}\right) & \text{for} \quad \theta > 0 \end{cases}$$

and plot it for $b = -a$.

5.9. Suppose that $\Theta = \{0, 1\}$, $D = \{d\} = [0, 1]$, and the loss function is $L(\theta, d) = |\theta - d|^a$, $a \geqslant 1$. Consider a class of decision functions of the form $\delta(x) \equiv \text{const}$ (i.e., the solution is found without

preliminary observations). Find in this class a Bayes solution for the a priori distribution of the parameter $\pi(0) = \alpha$, $\pi(1) = 1 - \alpha$, $\alpha \in [0, 1]$.

5.10*. (1) Show that for the risk of the Bayes solution in the classification problem (see Sec. 5.4) the representation

$$r(\delta^*) = \int \min_{1 \leqslant j \leqslant k} h_j(x)\, dx$$

is true. (For a discrete random variable below all the integrals are replaced by the respective sums.)

(2) Introduce the variables

$$I_{ij} = \int \min\, (\pi_i f_i(x),\ \pi_j f_j(x))\, dx,$$

$$\bar{I} = \max_{i \neq j} I(j|i),\ \underline{I} = \min_{i \neq j} I(j|i)$$

and prove the following estimates for $r(\delta^*)$:

$$\underline{I} \sum_{i=2}^{k} \max_{j<i} I_{ij} \leqslant r(\delta^*) \leqslant \bar{I} \sum_{1 \leqslant j < i \leqslant k} I_{ij}.$$

In what case do the estimates coincide?

Hint. Use the identity

$$\sum_{i=1}^{k} a_i = \max_{1 \leqslant j \leqslant k} a_j + \sum_{i=2}^{k} \min\, (a_i,\ \max_{j<i} a_j)$$

(use the induction on k in the proof).

5.11*. (Continued from Problem 5.10.) Let $F_i(\mathbf{x})$ be the distribution function of an r-variate non-degenerate normal law $\mathcal{N}(\boldsymbol{\mu}^{(i)}, \mathbf{A})$, $i = 1, 2$. Prove the formula

$$I_{12} = \pi_1 \Phi\left(-\frac{\sqrt{\varrho}}{2} - \frac{1}{\sqrt{\varrho}} \ln \frac{\pi_1}{\pi_2}\right) + \pi_2 \Phi\left(-\frac{\sqrt{\varrho}}{2} + \frac{1}{\sqrt{\varrho}} \ln \frac{\pi_1}{\pi_2}\right),$$

where $\varrho = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \mathbf{A}^{-1} (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$ is the *Mahalanobis distance* between the distributions $\mathcal{N}(\boldsymbol{\mu}^{(1)}, \mathbf{A})$ and $\mathcal{N}(\boldsymbol{\mu}^{(2)}, \mathbf{A})$. Derive a similar formula for I_{12} in the case of two Poisson's distributions.

5.12*. Construct the Bayes and minimax solutions for the classification problem with the two normal distributions given in the previous problem (compare with Problem 3.52).

5.13*. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathscr{L}(\xi) \in \mathscr{F} = \{F(x; \theta), \theta \in \Theta\}$, and let the a priori distribution of the parameter be $\mathscr{L}(\theta) \in \mathscr{F}^*$. The family $\mathscr{F}^*$ of the distributions of the parameter is said to be *conjugate* to $\mathscr{F}$ (denoted $\mathscr{F}^* \lhd \mathscr{F}$) if for $\mathbf{X} = \mathbf{x}$ the a posteriori distribution is $\mathscr{L}(\theta|\mathbf{x}) \in \mathscr{F}^*$.

Show that the following assertions are true $\left(\text{in what follows } x = \sum_{i=1}^{n} x_i\right)$:

(1) $B(a, b) \lhd \underline{Bi}(m, \theta)$ with $\mathscr{L}(\theta|\mathbf{x}) = B(a + x, b + nm - x)$.
(2) $B(a, b) \lhd \overline{Bi}(r, \theta)$ with $\mathscr{L}(\theta|\mathbf{x}) = B(a + x, b + nr)$.
(3) $\Gamma(a, \lambda) \lhd \Pi(\theta)$ with $\mathscr{L}(\theta|\mathbf{x}) = \Gamma\left(\frac{a}{na+1}, \lambda + x\right)$.

(4) $\Gamma(a, \lambda) \lhd \Gamma(\theta^{-1}, 1)$ with $\mathscr{L}(\theta|\mathbf{x}) = \Gamma\left(\frac{a}{ax+1}, \lambda + n\right)$.

(5) $\Pi(a, \alpha) \lhd R(0, \theta)$, where *Pareto's distribution* $\Pi(a, \alpha)$ is defined by the density $\pi(\theta) = \alpha a^{\alpha}/\theta^{\alpha+1}$, $\theta \geqslant a$ $(a, \alpha > 0)$, with $\mathscr{L}(\theta|\mathbf{x}) = \Pi(\max(a, x_1, \ldots, x_n), \alpha + n)$.

(6) $D(\alpha) \lhd M(n; \theta = (\theta_1, \ldots, \theta_N))$, where *Dirichlet's distribution* $D(\alpha)$, $\alpha = (\alpha_1, \ldots, \alpha_N)$, $\alpha_i > 0$, $i = 1, \ldots, N$, is defined by the density

$$\pi(\theta) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_N)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_N)} \theta_1^{\alpha_1 - 1} \ldots \theta_N^{\alpha_N - 1}, \quad \theta_1 + \ldots + \theta_N = 1,$$

with $\mathscr{L}(\theta|\mathbf{h} = (h_1, \ldots, h_N)) = D(\alpha + \mathbf{h})$.

(7) $\mathscr{N}(\mu, \sigma^2) \lhd \mathscr{N}(\theta, b^2)$ with $\mathscr{L}(\theta|\mathbf{x}) = \mathscr{N}(\mu_1, \sigma_1^2)$, where $\mu_1 = \sigma_1^2\left(\frac{\mu}{\sigma^2} + \frac{x}{b^2}\right)$, $\sigma_1^2 = \left(\frac{1}{\sigma^2} + \frac{n}{b^2}\right)^{-1}$.

Hint. It is sufficient to calculate the density of any distribution up to the normalizing factor and therefore, using the notation $f_\xi(t) = cp(t) \asymp p(t)$ for any random variable ξ (here the constant c is defined by the condition $c\int p(t)\,dt = 1$) when finding the a posteriori density $\pi(\theta|\mathbf{x}) = f(\mathbf{x}; \theta)\pi(\theta)/f(\mathbf{x})$, we may restrict ourselves to calculating the numerator $f(\mathbf{x}; \theta)\pi(\theta)$. We must act similarly when calculating the densities $\pi(\theta)$ and $f(\mathbf{x}; \theta)$.

5.14. Let us estimate the unknown probability θ of success from the given number of successes X in n Bernoulli trials (here

$\Theta = D = (0, 1)$). Suppose that the loss function has the form

$$L(\theta, d) = \frac{(d - \theta)^2}{\theta(1 - \theta)},$$

and the a priori distribution of the parameter $\mathscr{L}(\theta) = R(0, 1)$. Prove that the Bayes solution is $\delta^*(x) = x/n$ and that it is a minimax solution with $r(\delta^*) = 1/n$.

5.15. (Continued from Problem 5.14.) Find a Bayes solution δ^* for the case when the loss function is $L(\theta, d) = (d - \theta)^2$ and the a priori distribution is $\mathscr{L}(\theta) = B(a, b)$. For what values of the parameters a and b is δ^* a minimax solution? (Compare with Problem 2.6.)

| *Hint.* Use the solution to Problem 5.13 (1).

5.16. Consider the point estimation of a scalar parameter θ from the point of view of the decision theory when the decision set D coincides with the parametric set Θ and the decision $d \in D$ is the estimate of the parameter $\theta \in \Theta$. For a loss function of the form $L(\theta, d) = (d - \theta)^2$, the risk function $R(\theta, \delta) = \mathbf{E}_\theta(\delta(X) - \theta)^2$ is the mean square error of the estimate $\delta(X)$. Prove that the Bayes solution (Bayes estimate) $\delta^*(x)$ for the observation $X = x$ is

$$\delta^*(x) = \mathbf{E}(\theta | x) = \int \theta \pi(\theta | x)\, d\theta,$$

i.e., coincides with the a posteriori mean of the parameter, and the respective risk $r(\delta^*) = \mathbf{ED}(\theta | X)$, where

$$\mathbf{D}(\theta | x) = \mathbf{E}((\theta - \delta^*(x))^2 | x) = \int (\theta - \delta^*(x))^2 \pi(\theta | x)\, d\theta$$

is the variance of the a posteriori distribution of the parameter. The mathematical expectation is calculated with respect to the density $f(x)$ (probability in the discrete case). It is assumed here that all the respective moments do exist.

Apply the result to solve Problem 5.15.

5.17. Suppose that Bernoulli's trials are carried out until the rth failure, and X is the number of successes in these trials. Given an observation on X, construct a Bayes estimate for the unknown probability θ of success when the loss function is $L(\theta, d) = (d - \theta)^2$, and the a priori distribution is $\mathscr{L}(\theta) = B(a, b)$.

| *Hint.* Use the solutions to Problems 5.16 and 5.13 (2).

5.18. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from Poisson's distribution $\Pi(\theta)$, construct a Bayes estimate for the parameter if the loss function is $L(\theta, d) = (d - \theta)^2$, and the a priori distribution is $\mathscr{L}(\theta) = \Gamma(a, \lambda)$. Calculate the risk of this estimate and define the optimum sample size if the price of one observation is $c > 0$ (i.e., the size minimizing the total loss $r(\delta^*) + cn$).

[*Hint.* Use the solutions to Problems 5.16, 5.13 (3), and 1.39 (4).

5.19. (Continued from Problem 5.18.) Show that given the loss function $L(\theta, d) = (d - \theta)^2/\theta$, the Bayes estimate for $\lambda + \sum_{i=1}^{n} X_i > 1$ has the form

$$\delta^*(\mathbf{X}) = \frac{a}{na + 1}\left(\sum_{i=1}^{n} X_i + \lambda - 1\right),$$

and its risk is

$$r(\delta^*) = \frac{a}{na + 1}.$$

[*Hint.* When calculating the moments, use the formula $\Gamma(z + 1) = z\Gamma(z)$.

5.20. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$, estimate the parameter θ of an exponential distribution with the density $f(x; \theta) = \theta e^{-\theta x}$, $x > 0$. Suppose that the loss function is $L(\theta, d) = (1/\theta - d)^2$, and the a priori distribution is $\mathscr{L}(\theta) = \Gamma(a, \lambda)$, $\lambda > 2$. Prove that the Bayes estimate is of the form

$$\delta^*(\mathbf{X}) = \frac{1}{a(\lambda + n - 1)}\left(a\sum_{i=1}^{n} X_i + 1\right),$$

and its risk is

$$r(\delta^*) = (a^2(\lambda + n - 1)(\lambda - 1)(\lambda - 2))^{-1}.$$

Show that the optimal number of observations is

$$n^* = \frac{1}{a\sqrt{c(\lambda - 1)(\lambda - 2)}} - \lambda + 1$$

when the price of one observation is $c > 0$.

[*Hint.* Use the solutions to Problems 5.13 (4) and 1.39 (2).

5.21. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $R(0, \theta)$, where the a priori distribution of the parameter θ is Pareto's distribution with the parameters a and $\alpha > 2$ (see Problem 5.13 (5)). Show that the Bayes estimate for θ is of the form

$$\delta^*(\mathbf{x}) = \frac{n + \alpha}{n + \alpha - 1}\max(a, x_{(n)}), \quad x_{(n)} = \max_{1 \leqslant i \leqslant n} x_i,$$

and calculate its risk. Find the optimum sample size if the price of one observation is $c > 0$.

[*Hint.* Use the solutions to Problems 5.16, 5.13 (5), and 1.35.

5.22. Given an observation X, estimate the parameter θ of a uniform distribution $R(0, \theta)$, the parameter's a priori density being $\pi(\theta) = \theta e^{-\theta}$, $\theta > 0$. Prove that the Bayes estimate for a quadratic loss function has the form

$$\delta^*(X) = X + 1,$$

and its risk is

$$r(\delta^*) = 1.$$

Hint. Write the average a posteriori loss for the decision d in the form of an integral and differentiate it with respect to d. Use the formula $\Gamma(n+1) = \int_0^\infty t^n e^{-t}\, dt = n!$

5.23*. Let the vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ have a polynomial distribution $M(n; \boldsymbol{\theta} = (\theta_1, \ldots, \theta_N))$. Given an observation on $\boldsymbol{\nu}$, estimate $\boldsymbol{\theta}$ if the loss function is

$$L(\boldsymbol{\theta}, \mathbf{d}) = \sum_{i=1}^{N} (d_i - \theta_i)^2, \quad \mathbf{d} = (d_1, \ldots, d_N),$$

under the assumption that the parameter $\boldsymbol{\theta}$ has an a priori Dirichlet's distribution $D(\boldsymbol{\alpha})$ (see Problem 5.13 (6)). Show that for $\boldsymbol{\nu} = \mathbf{h} = (h_1, \ldots, h_N)$ the Bayes estimate $\boldsymbol{\delta}^*(\mathbf{h}) = (\delta_1^*(\mathbf{h}), \ldots, \delta_N^*(\mathbf{h}))$ has the form

$$\delta_i^*(\mathbf{h}) = \frac{\alpha_i + h_i}{\alpha + n}, \quad i = 1, \ldots, N, \quad \alpha = \sum_{i=1}^{N} \alpha_i,$$

and its risk is

$$r(\boldsymbol{\delta}^*) = \frac{\alpha^2 - \sum_{i=1}^{N} \alpha_i^2}{\alpha(\alpha+1)(\alpha+n)}.$$

Hint. Use the solutions to Problems 5.13 (6) and 1.52, and the formulas

$$\mathbf{E}\theta_i^r = \frac{\alpha_i(\alpha_i+1)\ldots(\alpha_i+r-1)}{\alpha(\alpha+1)\ldots(\alpha+r-1)}, \quad r = 1, 2, \ldots$$

for the moments of the distribution $D(\boldsymbol{\alpha})$.

5.24. Given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ from the normal distribution $\mathcal{N}(\theta, b^2)$, construct a Bayes estimate for the parameter, which will minimize the standard error for the a priori distribution $\mathcal{L}(\theta) = \mathcal{N}(\mu, \sigma^2)$. Calculate the risk of the resultant estimate and define the optimum number of observations if the price of one observation is $c > 0$.

Hint. Use the solutions to Problems 5.16 and 5.13 (7).

5.25*. Estimate a scalar parameter θ if the loss function is $L(\theta, d) = |\theta - d|$, $\theta, d \in R^1$.

(1) Prove that for $X = x$ the Bayes solution $d^* = \delta^*(x)$ is the median of the a posteriori distribution $\mathcal{L}(\theta|x)$ for any a priori distribution $\mathcal{L}(\theta)$, i.e., it is a number such that

$$\mathbf{P}(\theta \leqslant d^*|x) \geqslant \frac{1}{2}, \quad \mathbf{P}(\theta \geqslant d^*|x) \geqslant \frac{1}{2}.$$

(2) Use this result to estimate the mean in the model $\mathcal{N}(\theta, b^2)$ when $\mathcal{L}(\theta) = \mathcal{N}(\mu, \sigma^2)$.

Hints. (1) Prove the inequality $\mathbf{E}(|\theta - d||x) \geqslant \mathbf{E}(|\theta - d^*||x)$ $\forall d \in R^1$.

(2) Use the solution to Problem 5.13 (7).

CHAPTER 6

Statistics of Stationary Sequences

A sequence of random $\{X_t\}$, $t = \ldots, -1, 0, 1, \ldots$, which is unlimited on both ends is said to be *stationary* if the conditions

$$\mathbf{E}X_t = m = \text{const},$$
$$\operatorname{cov}(X_{k+t}, X_t) = \mathbf{E}(X_{k+t} - m)(X_t - m) = R_k$$

are met.

A sequence of numbers $\{R_k\}$, $k = \ldots, -1, 0, 1, \ldots$, is called the *covariance function* of the sequence $\{X_t\}$. Here $R_{-k} = R_k$ for all k and $R_0 = \mathbf{D}X_t = \sigma^2 = \text{const}$. We will assume that

$$\sum_{k=1}^{\infty} |R_k| < \infty. \tag{6.1}$$

The statistics

$$\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t, \quad \tilde{C}_k(n) = \frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - \overline{X})(X_{k+t} - \overline{X}),$$
$$k = 0, 1, \ldots, n-1,$$

are used to estimate m and R_k from the observations $X_1, \ldots, X_n$.

As an example we simulate n terms of a stationary sequence

$$X_t = \xi_{t-1} + \xi_t + \xi_{t+1}, \quad t = 1, 2, \ldots, n, \tag{6.2}$$

where ξ_t, $t = 0, \pm1, \pm2, \ldots$, are independent random variables uniformly distributed on the segment $[0, 2h]$. We can easily show that $\mathbf{E}X_t = 3h$, $R_0 = h/2$, $R_1 = h/3$, $R_2 = h/6$, $R_i = 0$, $i \geqslant 3$. Tables 6.1 and 6.2 give the values of the statistics $\overline{X}$ and $\tilde{C}_k(n)$ for different n.

Table 6.1

n	$\overline{X}$	$\tilde{C}_0(n)$	$\tilde{C}_1(n)$	$\tilde{C}_2(n)$	$\tilde{C}_3(n)$	$\tilde{C}_4(n)$
10	1.70	0.372	0.264	0.088	-0.128	-0.242
100	1.41	0.270	0.185	0.092	0.026	0.047
1000	1.50	0.262	0.178	0.089	0.002	-1.0×10^{-4}

$h = 0.5$ ($\mathrm{E}X_t = 1.5$, $R_0 = 0.25$, $R_1 = 0.166\ldots$, $R_2 = 0.0833\ldots$).

Table 6.2

n	$\overline{X}$	$\tilde{C}_0(n)$	$\tilde{C}_1(n)$	$\tilde{C}_2(n)$	$\tilde{C}_3(n)$	$\tilde{C}_4(n)$
10	2.040	0.535	0.381	0.127	-0.184	-0.349
100	1.690	0.389	0.266	0.132	0.039	0.068
1000	1.800	0.378	0.257	0.126	0.003	-2.0×10^{-4}

$h = 0.6$ ($\mathrm{E}X_t = 1.8$, $R_0 = 0.3$, $R_1 = 0.2$, $R_2 = 0.1$).

The *spectral density* $f(x)$ is an important characteristic of a stationary sequence $\{X_t\}$. It is a Fourier transform of the covariance function $\{R_k\}$, i.e.,

$$f(x) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R_k \cos kx, \ x \in [-\pi, \pi]. \tag{6.3}$$

The spectral density (if it exists) and the covariance function are in a one-to-one correspondence. The statistics of the form

$$\tilde{f}_n(x) = \frac{1}{2\pi} \sum_{|k| \leqslant n-1} w_n(k)\tilde{C}_k(n) \cos kx \tag{6.4}$$

are used to estimate $f(x)$ from the observations $X_1, \ldots, X_n$, where $\{w_n(k)\}$ is a sequence of weighting coefficients ($w_n(-k) = w_n(k)$). Specifically, for $w_n(k) = 1 - (|k|/n)$ we obtain the *periodogram* of the sample. If the mean $m = \mathbf{E}X_t$ is known, then we replace $\overline{X}$ by m in (6.4).

Problems

6.1. Prove that the arithmetic mean $\overline{X} = (X_1 + \ldots + X_n)/n$ is an unbiased and consistent estimator for $m = \mathbf{E}X_t$.

6.2. Prove that the statistic

$$\tilde{C}_k(n) = \frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - m)(X_{k+t} - m), \quad 0 \leqslant k < n,$$

is an unbiased estimator for R_k.

6.3*. Prove that the statistic

$$\bar{C}_k(n) = \frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - \overline{X})(X_{k+t} - \overline{X})$$

is an asymptotically unbiased estimator for R_k as $n \to \infty$, i.e., $\mathbf{E}\bar{C}_k(n) \to R_k$ (k is fixed).

6.4. Let ξ and η be random variables with $\mathbf{E}\xi = \mathbf{E}\eta = 0$, $\mathbf{D}\xi = \mathbf{D}\eta = \sigma^2$, $\operatorname{cov}(\xi, \eta) = 0$. Prove that the sequence $X_t = \xi \cos \lambda t + \eta \sin \lambda t$, $t = 0, \pm 1, \pm 2, \ldots$, $\lambda \in (0, \pi)$, is stationary and calculate its covariance function.

6.5. Let ξ_t, $t = 0, \pm 1, \pm 2, \ldots$, be uncorrelated random variables with $m = \mathbf{E}\xi_t$, $\sigma^2 = \mathbf{D}\xi_t$. Is the sequence $\{\xi_t\}$ stationary?

Prove that the sequence

$$X_t = \sum_{j=0}^{r} \alpha_j \xi_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots,$$

is stationary. Find $\mathbf{E}X_t$ and R_k.

6.6*. Prove that for the sequence (6.2) the estimator $\tilde{C}_k(n)$ found in Problem 6.2 is consistent.

6.7. Simulate a sequence of the form (6.2) for the case when ξ_t are normally distributed with $\mathbf{E}\xi_t = 0.5$, $\mathbf{D}\xi_t = 0.1$, and $n = 100$. Compile the respective table similar to Table 6.1.

6.8*. Given the values X_t, $t = -n, -n+1, \ldots, -1, 0$, predict the value of X_1, i.e., find the *optimal linear predictor* $X_{1n}^* = \sum_{t=-n}^{0} \beta_{tn}^* X_t$, which means that β_t must minimize the expression $\mathbf{E}\left(X_1 - \sum_{t=-n}^{0} \beta_t X_t\right)^2$. Calculate the *minimal mean square error* $\sigma^2(n) = \mathbf{E}(X_1 - X_{1n}^*)^2$ *of the prediction.*

6.9. Let ν_t, $t = 0, \pm 1, \pm 2, \ldots$, be a stationary *Markov chain* with states 1 and 2 [2].

Prove that the matrix $\|p_{ij}(t)\|_1^2$, where

$$p_{ij}(t) = \mathbf{P}(\nu_{s+t} = j \mid \nu_s = i), \quad i, j = 1, 2,$$

is defined by the formula

$$\|p_{ij}(t)\| = \frac{1}{2}\left(\begin{Vmatrix} 1 & 1 \\ 1 & 1 \end{Vmatrix} + (1 - 2\alpha)^t \begin{Vmatrix} 1 & -1 \\ -1 & 1 \end{Vmatrix}\right)$$

if

$$\|p_{ij}(1)\| = \begin{Vmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{Vmatrix}, \quad 0 < \alpha < 1.$$

Find a stationary distribution for this chain.

6.10. Make a program for simulating the sequence ν_t, $t = 0, 1, \ldots, n$, defined in Problem 6.9.

6.11. Let $\{\nu_t\}$ be the stationary Markov chain defined in Problem 6.9. Is the sequence $\{\eta_t\}$ with

$$\eta_t = \begin{cases} 1 & \text{for} \quad \nu_t = 2, \\ -1 & \text{for} \quad \nu_t = 1 \end{cases}$$

stationary? Find $\mathbf{E}\eta_t$ and R_k.

6.12. Simulate a sequence $\eta_1, \ldots, \eta_{100}$, where η_t is defined as in the previous problem and $\alpha = 1/3$. Calculate the estimates for $\mathbf{E}\eta_t$ and R_k.

6.13. Let ν_t be the Markov chain from Problem 6.9, and let $\xi_1(t)$, $\xi_2(t)$, $t = 0, \pm 1, \pm 2, \ldots$, be independent stationary sequences with $\mathbf{E}\xi_i(t) = 0$ and the covariance functions $R_k^{(i)}$, $i = 1, 2$. We assume that $\eta_t = \xi_{\nu_t}(t)$. Is the sequence $\{\eta_t\}$ stationary? Find $\mathbf{E}\eta_t$ and R_k.

Hint. Use the formula for a total expectation.

6.14. Simulate a sequence $\eta_1, \ldots, \eta_{100}$, where η_t is defined as in Problem 6.13, $\alpha = 1/3$, and $\xi_i(t)$ are uniformly distributed on the segment $[-1, 1]$. Calculate the estimates for $\mathbf{E}\eta_t$ and R_k.

6.15. Solve Problem 6.14 for $\mathscr{L}(\xi_i(t)) = \mathscr{N}(0, 1)$.

6.16. Show that under the condition (6.1) the spectral density $f(x)$ (see (6.3)) exists, is continuous, and defines the covariance function according to the formula

$$R_k = \int_{-\pi}^{\pi} f(x) \cos kx \, dx.$$

6.17. Calculate the spectral density for a stationary sequence of uncorrelated random variables.

6.18. Does the spectral density of the sequence $\{X_t\}$ in Problem 6.4 exist? Show that in this case $R_k = \int_{-\pi}^{\pi} \cos kx \, dF(x)$, where $F(x)$ is a step function with steps $\sigma^2/2$, $F(-\pi) = 0$, $F(\pi) = \sigma^2$ at the points $\pm\lambda$. ($F(x)$ is called the *spectral function* of the sequence $\{X_t\}$.)

6.19. Construct the following representation of the periodogram (see (6.4)) using the sample values of the sequence $\{X_t\}$ (the mean m is known):

$$\tilde{f}_n(x) = \frac{n}{8\pi} R_n^2(x), \quad R_n^2(x) = A_n^2(x) + B_n^2(x),$$

where

$$\left.\begin{matrix} A_n(x) \\ B_n(x) \end{matrix}\right\} = \frac{2}{n} \sum_{t=1}^{n} (X_t - m) \begin{cases} \cos xt, \\ \sin xt. \end{cases}$$

6.20*. Show that for the expectation of a periodogram with the known mean $m = \mathbf{E}X_t$ the representation

$$\mathbf{E}\tilde{f}_n(x) = \int_{-\pi}^{\pi} k_n(x-y) f(y) \, dy$$

is true, where $k_n(x) = \dfrac{1}{2\pi n}\left[\sin\left(\dfrac{nx}{2}\right) \Big/ \sin\left(\dfrac{x}{2}\right)\right]^2$ is *Fejér's kernel*.

Prove that $\mathbf{E}\tilde{f}_n(x) \to f(x)$, $-\pi \leqslant x \leqslant \pi$ as $n \to \infty$.

Remark. A periodogram is an asymptotically unbiased estimator for the spectral density even if m is unknown but the estimates are not consistent. Consistent estimates can be obtained if the weighting coefficients $w_n(k)$ in (6.4) are chosen properly.

6.21*. Let the mean $m = \mathbf{E}X_t$ be known and $w_n(k) = \left(1 - \dfrac{|k|}{n}\right) \times \dfrac{\sin \varepsilon k}{\varepsilon k}$. Prove that $\tilde{f}_n(k)$ is an asymptotically unbiased and consis-

tent estimator for the quantity $\frac{1}{2\varepsilon}\int_{\lambda-\varepsilon}^{\lambda+\varepsilon} f(y)\,dy$, $0 < \varepsilon < \pi$, $\lambda \in [-\pi + \varepsilon, \pi - \varepsilon]$.

Remark. The result is also true for an unknown m.

6.22*. Let m be known, and $w_n(k) = \left(1 - \frac{|k|}{n}\right)\left(1 - \frac{|k|}{l_n}\right)$ for $|k| \leqslant l_n$ and $w_n(k) = 0$ for $|k| > l_n$. Prove that if $n, l_n \to \infty$, $l_n/n \to 0$, and $\sum_k |kR_k| < \infty$, then $\tilde{f}_n(x)$ is an asymptotically unbiased estimator for $f(x)$.

Remark. The result is also true for an unknown m. The estimates are consistent under wide conditions.

ANSWERS AND SOLUTIONS

TO CHAPTER 1

1.1. Take $X_n = 1$ if $U_n \leqslant p$ and $X_n = 0$ if $U_n > p$, where U_n is the sequence (1.5).

1.3. Divide the segment $[0, 1]$ into N parts $\Delta_1, \Delta_2, \ldots, \Delta_N$, where $\Delta_1 = [0, p_1]$, $\Delta_l = [p_1 + \ldots + p_{l-1}, p_1 + \ldots + p_l]$, $l = 2, 3, \ldots, N$. Take $X_k = l$ if $U_k \in \Delta_l$, $l = 1, \ldots, N$, where U_n is the sequence (1.5). The numbers $X_1, \ldots, X_n$ form the realization of n first trials of the polynomial scheme with the indicated probabilities of the outcomes.

1.4. Assume that $S_0 = 0$, $S_n = S_{n-1} + X_n$, $n \geqslant 1$, where $X_n = 1$ if $U_n \leqslant 1/2$ and $X_n = -1$ if $U_n > 1/2$. Here U_n is the sequence (1.5).

1.6. Suppose that $X_n = -a \ln (1 - U_n)$, where U_n is the sequence (1.5).

1.8. Let $\xi_1, \ldots, \xi_m$ be independent random variables having an exponential distribution with the parameter a. Then the random variable $\eta = \xi_1 + \ldots + \xi_m$ has Erlang's distribution with the parameters (a, m).

1.9. Take

$$X_n = \frac{U_{(n-1)N+1} + U_{(n-1)N+2} + \ldots + U_{nN} - (1/2)N}{\sqrt{N/12}},$$

where U_n is the sequence (1.5). We can obtain a good approximation to the normal distribution already for $N = 12$. Take this value for the calculations.

1.12. If $\xi_1, \ldots, \xi_k$ are independent Bernoulli random variables with the parameter p (see Problem 1.1), then $\mathscr{L}(\xi_1 + \ldots + \xi_k) = Bi(k, p)$.

1.13. By the De Moivre-Laplace theorem we have for large n

$$\mathbf{P}\left(\frac{|\nu_n - np|}{\sqrt{npq}} \leqslant t\right) = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1, \quad q = 1 - p,$$

or

$$\mathbf{P}\left(\left|\frac{\nu_n}{n} - p\right| \leqslant t\sqrt{\frac{pq}{n}}\right) = 2\Phi(t) - 1.$$

Consequently, we must take $\delta_\gamma = t\sqrt{\dfrac{pq}{n}}$ and $t = \Phi^{-1}\left(\dfrac{1 + \gamma}{2}\right) \equiv u_{(1+\gamma)/2}$ for the relation $\mathbf{P}\left(\left|\dfrac{\nu_n}{n} - p\right| \leqslant \delta_\gamma\right) = \gamma$ to be true. For $\gamma = 0.98$ the quantile is

$u_{0.99} = 2.326$. For the given experimental data the boundary is $\delta_{0.98} = 0.0183$, and $\left|\frac{h}{n} - \frac{1}{2}\right| = 0.0069$. This is in good agreement with the theory.

1.14. According to our assumption, the appearance of a number no greater than 4 can be considered to be a success in the Bernoulli trials with the probability $p = 1/2$ of success. Therefore (see the solution to Problem 1.13), the boundary is $\delta_{0.98} = 0.0116$, and the observed deviation of the success frequency is $\left|\frac{h}{n} - \frac{1}{2}\right| = 0.0089 < \delta_{0.98}$. Consequently, we have a good agreement between the experimental data and the theory.

1.16. For large n we have

$$\mathbf{P}\left(\sqrt{\frac{n}{\mu_2}}\,|\overline{X} - \alpha_1| \leqslant t\right) \approx 2\Phi(t) - 1,$$

or

$$\mathbf{P}(|\overline{X} - \alpha_1| \leqslant \delta) \approx 2\Phi\left(\sqrt{\frac{n}{\mu_2}}\,\delta\right) - 1.$$

In this case $\alpha_1 = 6$, $\mu_2 = 3$, $n = 4096$, and the right-hand side of the approximate equation is equal to 0.998 for $\delta = \sqrt{\mu_2/n}\,u_{0.999} = \sqrt{3/4096} \times 3.090 = 0.0836\ldots$. The observed value $|\overline{x} - 6| = 0.1389$ is much greater than the boundary, i.e., we have observed a hardly probable event.

1.17. In this case (see the previous solution) $\alpha_1 = 2$, $\mu_2 = 5/6$, $\delta = \sqrt{5/(6 \times 4096)} \times 3.090 = 0.044$. The observed value $|\overline{x} - 2| = 0.003$ lies within these boundaries, i.e., this characteristic gives a better agreement of the data with the theory.

1.19. Using the results of Problems 1.13 and 1.16, we obtain $\delta = \sqrt{\mu_2/n}\,u_{0.99} = \sqrt{11.9167/500} \times 2.326 = 0.359$. The observed deviation is $|\overline{x} - \alpha_1| = |5.942 - 6| = 0.058$, i.e., we have a good agreement of the data with the theory.

1.24. Since $F_n(x_0) = \frac{\mu_n(x_0)}{n}$ and $\mathscr{L}(\mu_n(x_0)) = Bi(n, p_0)$, where $p_0 = F(x_0)$, using the De Moivre-Laplace theorem for $n \to \infty$, we have

$$\mathbf{P}\left(\frac{\mu_n(x_0) - np_0}{\sqrt{np_0q_0}} \leqslant z\right) = \Phi(z), \quad q_0 = 1 - p_0.$$

Whence

$$\mathbf{P}\left(|F_n(x_0) - p_0| \leqslant \frac{t}{\sqrt{n}}\right) \to \Phi\left(\frac{t}{\sqrt{p_0q_0}}\right) - \Phi\left(-\frac{t}{\sqrt{p_0q_0}}\right)$$

$$= 2\Phi\left(\frac{t}{\sqrt{p_0q_0}}\right) - 1.$$

1.25. We have

$$\operatorname{cov}(F_n(x_1), F_n(x_2)) = \frac{1}{n^2}\operatorname{cov}(\mu_n(x_1), \Delta_n(x_1, x_2) + \mu_n(x_1))$$

$$= \frac{1}{n^2}[\operatorname{cov}(\mu_n(x_1), \Delta_n(x_1, x_2)) + \mathbf{D}\mu_n(x_1)].$$

Here $\mathbf{D}\mu_n(x_1) = nF(x_1)(1 - F(x_1))$ (see the solution to Problem 1.24), and

$$\operatorname{cov}(\mu_n(x_1), \Delta_n(x_1, x_2)) = \sum_{i, j=1}^{n} \operatorname{cov}(\eta_i, \zeta_j).$$

Since the observations are independent, the indicators η_i and ζ_j are independent for $i \neq j$, and we obtain $\operatorname{cov}(\eta_i, \zeta_j) = 0$. We also get

$$\operatorname{cov}(\eta_i, \zeta_i) = \mathbf{E}\eta_i\zeta_i - \mathbf{E}\eta_i\mathbf{E}\zeta_i = \mathbf{P}(\eta_i = \zeta_i = 1)$$

$$-\mathbf{P}(\eta_i = 1)\mathbf{P}(\zeta_i = 1) = -\mathbf{P}(\eta_i = 1)\mathbf{P}(\zeta_i = 1) = -F(x_1)(F(x_2) - F(x_1)),$$

because $\{\eta_i = \zeta_i = 1\}$ is an impossible event. From this we find

$$\operatorname{cov}(\mu_n(x_1), \Delta_n(x_1, x_2)) = -nF(x_1)(F(x_2) - F(x_1)).$$

By uniting these formulas, we get the desired result.

1.26. Consider a complete group of N events $\mathbf{E}_1 = \{\xi \leqslant x_1\}$, $\mathbf{E}_2 = \{x_1 < \xi \leqslant x_2\}, \ldots, \mathbf{E}_{N-1} = \{x_{N-2} < \xi \leqslant x_{N-1}\}$, $\mathbf{E}_N = \{\xi > x_{N-1}\}$, whose probabilities are $p_1, \ldots, p_N$, respectively. Then ν_i is obviously the number of the realizations of $\mathbf{E}_i$ in n independent and uniform trials, $i = 1, \ldots, N$. Consequently, $\mathscr{L}(\boldsymbol{\nu}) = M(n; p_1, \ldots, p_N)$. We have

$$-np_1p_2 = \operatorname{cov}(\nu_1, \nu_2) = \operatorname{cov}(\mu_n(x_1), \mu_n(x_2) - \mu_n(x_1))$$

$$= \operatorname{cov}(\mu_n(x_1), \mu_n(x_2)) - \mathbf{D}\mu_n(x_1).$$

Whence

$$\operatorname{cov}(\mu_n(x_1), \mu_n(x_2)) = \mathbf{D}\mu_n(x_1) - np_1p_2 = np_1(1 - p_1 - p_2)$$

$$= nF(x_1)(1 - F(x_2)),$$

which is equivalent to the result obtained in Problem 1.25.

1.27. The random variables X_i^k, $i = 1, \ldots, n$, are independent and distributed as ξ^k for any k. Therefore,

$$\operatorname{cov}(A_{nk}, A_{ns}) = \frac{1}{n^2}\sum_{i, j=1}^{n}\operatorname{cov}(X_i^k, X_j^s) = \frac{1}{n}\operatorname{cov}(\xi^k, \xi^s)$$

$$= \frac{1}{n}(\mathbf{E}\xi^{k+s} - \mathbf{E}\xi^k\mathbf{E}\xi^s) = \frac{1}{n}(\alpha_{k+s} - \alpha_k\alpha_s).$$

In particular,

$$\mathbf{D}\overline{X} = \mathbf{D}A_{n1} = \frac{1}{n}(\alpha_2 - \alpha_1^2) = \frac{\mu_2}{n}.$$

When investigating the central sampling moments, we may assume that $\alpha_1 = 0$ and hence $\mu_k = \alpha_k$. Taking this into account, we have

$$\mathbf{E}S^2 = \mathbf{E}A_{n2} - \mathbf{E}A_{n1}^2 = \mu_2 - \frac{\mu_2}{n} = \frac{n-1}{n}\mu_2.$$

We know that $(S^2)^2 = A_{n2}^2 - 2A_{n1}^2A_{n2} + A_{n1}^4$ and directly calculate

$$\mathbf{E}A_{n2}^2 = \frac{1}{n^2}\mathbf{E}\left(\sum_{i=1}^{n} X_i^4 + \sum_{i\neq j} X_i^2X_j^2\right) = \frac{1}{n^2}(n\mu_4 + n(n-1)\mu_2^2)$$

$$= \frac{\mu_4 + (n-1)\mu_2^2}{n},$$

$$\mathbf{E}A_{n1}^2A_{n2} = \frac{1}{n^3}\mathbf{E}\left(\sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_iX_j\right)\sum_{k=1}^{n} X_k^2$$

$$= \frac{1}{n^3}\mathbf{E}\left(\sum_{i=1}^{n} X_i^2\right)^2 = \frac{\mu_4 + (n-1)\mu_2^2}{n^2},$$

$$\mathbf{E}A_{n1}^4 = \frac{1}{n^4}\mathbf{E}\left(\sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_iX_j\right)^2 = \frac{1}{n^4}\mathbf{E}\left(\left(\sum_{i=1}^{n} X_i^2\right)^2 + 2\sum_{i\neq j} X_i^2X_j^2\right)$$

$$= \frac{\mu_4 + (n-1)\mu_2^2}{n^3} + \frac{2(n-1)}{n^3}\mu_2^2 = \frac{\mu_4 + 3(n-1)\mu_2^2}{n^3}.$$

Whence

$$\mathbf{D}S^2 = \mathbf{E}(S^2)^2 - (\mathbf{E}S^2)^2 = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3},$$

which is equivalent to the given formula. We still consider that $\alpha_1 = 0$ and get $\operatorname{cov}(\overline{X}, S^2) = \mathbf{E}(\overline{X}S^2)$. By writing

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left(\sum_{j=1}^{n} X_j\right)^2 = \frac{n-1}{n^2}\sum_{j=1}^{n} X_j^2 - \frac{1}{n^2}\sum_{i\neq j} X_iX_j,$$

we obtain

$$\mathbf{E}(\overline{X}S^2) = \frac{n-1}{n^3}\mathbf{E}\left(\sum_{i=1}^{n}X_i\right)\left(\sum_{j=1}^{n}X_j^2\right)$$

$$= \frac{n-1}{n^3}\mathbf{E}\left(\sum_{i=1}^{n}X_i^3\right) = \frac{n-1}{n^2}\mu_3.$$

For the distribution $\mathcal{N}(\mu, \sigma^2)$ the moments are $\alpha_1 = \mu$, $\mu_2 = \sigma^2$, $\mu_3 = 0$, $\mu_4 = 3\sigma^4$. Therefore, we have $\mathbf{E}\overline{X} = \mu$, $\mathbf{D}\overline{X} = \frac{\sigma^2}{n}$, $\mathbf{E}S^2 = \frac{n-1}{n}\sigma^2$, $\mathbf{D}S^2 = \frac{2(n-1)}{n^2}\sigma^4$, $\operatorname{cov}(\overline{X}, S^2) = 0$.

1.28. Consider the r-dimensional vectors $\xi_s = (X_s^{k_1}, \ldots, X_s^{k_r})$, $s = 1, \ldots, n$. They are independent and similarly distributed with $\mathbf{E}(\xi_1) = \alpha$, $\mathbf{D}(\xi_1) = \|\operatorname{cov}(X_1^{k_i}, X_1^{k_j})\|_1^r = \Sigma$ (α and Σ are defined in the statement of the problem). Using the Central Limit Theorem, we then have for $n \to \infty$

$$\mathcal{L}\left(\frac{1}{\sqrt{n}}(\xi_1 + \ldots + \xi_n - n\alpha)\right) \to \mathcal{N}(0, \Sigma).$$

It remains to recall that $\frac{1}{\sqrt{n}}(\xi_1 + \ldots + \xi_n - n\alpha) = \sqrt{n}(A_{nk_1} - \alpha_{k_1}, \ldots, A_{nk_r} - \alpha_{k_r})$.

1.29. We can consider that $\alpha_1 = \mathbf{E}\xi = 0$ (see the solution to Problem 1.27). We take $\eta_n = \sqrt{n}(S_n^2 - \mu_2) = \xi_n + \delta_n$, where $\xi_n = \sqrt{n}(A_{n2} - \mu_2)$, $\delta_n = -\sqrt{n}A_{n1}^2$. Since we have assumed that $\mathcal{L}(\xi_n) \to \mathcal{N}(0, \mu_4 - \mu_2^2)$ (see the solution to Problem 1.28), it is sufficient to show that $\delta_n \xrightarrow{\mathbf{P}} 0$. But

$$\mathbf{P}(|\delta_n| > \varepsilon) \leqslant \frac{1}{\varepsilon}\mathbf{E}|\delta_n| = \frac{\sqrt{n}}{\varepsilon}\mathbf{E}A_{n1}^2 = \frac{\sqrt{n}}{\varepsilon}\mathbf{D}A_{n1} = \frac{\mu_2}{\varepsilon\sqrt{n}} \to 0,$$

which was to be proved. The asymptotics of the moments follows from Problem 1.27.

1.30. The events $\{X_{(r)} \leqslant x_1, X_{(s)} \leqslant x_2\}$ and $\{\mu_n(x_1) \geqslant r, \mu_n(x_2) \geqslant s\}$ are equivalent. Therefore, we have $F_{rs}(x_1, x_2) = \mathbf{P}(\mu_n(x_1) \geqslant r, \mu_n(x_2) \geqslant s)$. Suppose that $x_1 < x_2$. We consider the random variables $\nu_1 = \mu_n(x_1)$, $\nu_2 = \mu_n(x_2) - \mu_n(x_1)$, $\nu_3 = n - \mu_n(x_2)$. Then (see the solution to Problem 1.26)

$$\mathcal{L}(\nu_1, \nu_2, \nu_3) = M(n; p_1, p_2, p_3),$$

where $p_1 = F(x_1)$, $p_2 = F(x_2) - F(x_1)$, $p_3 = 1 - F(x_2)$. From this we have

$$\mathbf{P}(\mu_n(x_1) \geqslant r, \quad \mu_n(x_2) \geqslant s) = \Sigma\mathbf{P}(\nu_1 = m, \nu_2 = j),$$

where the summation is performed over all m and j, which satisfy the conditions $m \geqslant r$, $s \leqslant m + j \leqslant n$. Since

$$\mathbf{P}(\nu_1 = m, \nu_2 = j) = \frac{n!}{m!j!(n-m-j)!} p_1^m p_2^j p_3^{n-m-j},$$

we obtain the formula from the statement of the problem. If $x_1 \geqslant x_2$, we have $\{X_{(r)} \leqslant x_1, X_{(s)} \leqslant x_2\} = \{X_{(s)} \leqslant x_2\}$. The formula for the univariate distribution function can be obtained, for example, from $F_r(x_1) = \lim\limits_{x_2 \to \infty} F_{rs}(x_1, x_2)$.

1.31. Suppose that $r = 2$ (the general case is treated in a similar way) and the points $x_1 < x_2$ are given. The event $\{X_{(k_1)} \in (x_1; x_1 + dx_1), X_{(k_2)} \in (x_2; x_2 + dx_2)\}$ occurs if and only if $k_1 - 1$ of all the observations are smaller than x_1, one observation is in the interval $(x_1, x_1 + dx_1)$, $k_2 - k_1 - 1$ observations are in the interval $(x_1 + dx_1, x_2)$, one observation is in the interval $(x_2, x_2 + dx_2)$, and the other $n - k_2$ observations are greater than $x_2 + dx_2$. Since the observations are independent, the probability of this event for small dx_1 and dx_2, up to the terms with a greater order of smallness, is equal to

$$C_n^{k_1-1} F^{k_1-1}(x_1)(n - k_1 + 1) f(x_1)\, dx_1$$
$$\times\, C_{n-k_1}^{k_2-k_1-1} (F(x_2) - F(x_1))^{k_2-k_1-1} (n - k_2 + 1) f(x_2)\, dx_2 (1 - F(x_2))^{n-k_2}.$$

We divide this by $dx_1 dx_2$ and tend dx_1 and dx_2 to zero to obtain the desired formula for $g_{k_1k_2}(x_1, x_2)$.

1.32. We write $k_i = [np_i]$, $i = 1, 2$, and assume that $\eta_{ni} = (Z_{np_i} - \zeta_{p_i})\sqrt{n}$, $i = 1, 2$. By (1.2) the joint distribution density of the random variables η_{n_1} and η_{n_2} is (see Problem 1.31)

$$\varphi_n(y_1, y_2) = \frac{1}{n} g_{k_1+1, k_2+1}\left(\zeta_{p_1} + \frac{y_1}{\sqrt{n}}, \zeta_{p_2} + \frac{y_2}{\sqrt{n}}\right) = A_1(n) A_2(n) A_3(n),$$

where

$$A_1(n) = \frac{n! p_1^{k_1} (p_2 - p_1)^{k_2-k_1-1} (1 - p_2)^{n-k_2-1}}{n k_1! (k_2 - k_1 - 1)! (n - k_2 - 1)!},$$

$$A_2(n) = f\left(\zeta_{p_1} + \frac{y_1}{\sqrt{n}}\right) f\left(\zeta_{p_2} + \frac{y_2}{\sqrt{n}}\right),$$

$$A_3(n) = \left(\frac{1}{p_1} F\left(\zeta_{p_1} + \frac{y_1}{\sqrt{n}}\right)\right)^{k_1} \left(\frac{F\left(\zeta_{p_2} + \frac{y_2}{\sqrt{n}}\right) - F\left(\zeta_{p_1} + \frac{y_1}{\sqrt{n}}\right)}{p_2 - p_1}\right)^{k_2-k_1-1}$$
$$\times \left(\frac{1 - F\left(\zeta_{p_2} + \frac{y_2}{\sqrt{n}}\right)}{1 - p_2}\right)^{n-k_2-1}.$$

From Stirling's formula it follows that $A_1(n) \to \dfrac{1}{2\pi\sqrt{p_1(p_2 - p_1)(1 - p_2)}}$. Since the distribution density $f(x)$ is continuous, we have $A_2(n) \to f(\zeta_{p_1})f(\zeta_{p_2})$. Finally, since

$$F\left(\zeta_{p_i} + \frac{y_i}{\sqrt{n}}\right) = p_i + f(\zeta_{p_i})\frac{y_i}{\sqrt{n}} + f'(\zeta_{p_i})\frac{y_i^2}{2n} + o\left(\frac{1}{n}\right), \quad i = 1, 2,$$

we easily get

$$\ln A_3(n) \to -\frac{1}{2}\left(\frac{p_2}{p_1(p_2 - p_1)}f^2(\zeta_{p_1})y_1^2 - \frac{2y_1y_2}{p_2 - p_1}f(\zeta_{p_1})f(\zeta_{p_2})\right.$$
$$\left. + \frac{(1 - p_1)y_2^2}{(1 - p_2)(p_2 - p_1)}f^2(\zeta_{p_2})\right) = -\frac{1}{2}\sum_{i,j=1}^{2}\sigma^{ij}y_iy_j, \quad |\sigma^{ij}| = |\sigma_{ij}|^{-1}.$$

Taking into account that $f(\zeta_{p_1})f(\zeta_{p_2})/\sqrt{p_1(p_2 - p_1)(1 - p_2)} = (\det|\sigma_{ij}|)^{-1/2}$, we finally write

$$\varphi_n(y_1, y_2) \to \frac{1}{2\pi\sqrt{\det|\sigma_{ij}|}}\exp\left\{-\frac{1}{2}\sum_{i,j=1}^{2}\sigma^{ij}y_iy_j\right\},$$

i.e., in the limit we have the density of a bivariate normal distribution with zero means and the matrix of the second moments $|\sigma_{ij}|$.

1.33. According to the solution of Problem 1.31, the joint distribution density of $X_{(r)}$ and $X_{(n-s+1)}$ is (for $x_1 < x_2$)

$$g_{r,n-s+1}(x_1, x_2) = \frac{n!}{(r-1)!(s-1)!(n-s-r)!}F^{r-1}(x_1)[F(x_2) - F(x_1)]^{n-s-r}$$
$$\times [1 - F(x_2)]^{s-1}f(x_1)f(x_2).$$

Since the Jacobian of the transformation $y_1 = nF(x_1)$, $y_2 = n[1 - F(x_2)]$ is $J(x_1, x_2) = -n^2f(x_1)f(x_2)$, by formula (1.2) the joint distribution density of the random variables $\varkappa_n = nF(X_{(r)})$ and $\eta_n = n[1 - F(X_{(n-s+1)})]$ has the form

$$\varphi_n(y_1, y_2) = g_{r,n-s+1}\left(F^{-1}\left(\frac{y_1}{n}\right), F^{-1}\left(1 - \frac{y_2}{n}\right)\right)$$

$$\Bigg/\left|J\left(F^{-1}\left(\frac{y_1}{n}\right), F^{-1}\left(1 - \frac{y_2}{n}\right)\right)\right| = \frac{n!}{(n-r-s)!n^{r+s}}\frac{y_1^{r-1}}{(r-1)!}\frac{y_2^{s-1}}{(s-1)!}$$

$$\times\left(1 - \frac{y_1 + y_2}{n}\right)^{n-r-s} \to \frac{y_1^{r-1}}{(r-1)!}e^{-y_1}\frac{y_2^{s-1}}{(s-1)!}e^{-y_2}$$

if $n \to \infty$ and r, s are fixed.

Thus x_n and η_n are asymptotically independent and so are $X_{(r)}$ and $X_{(n-s+1)}$. We also have $\mathscr{L}(x_n) \to \Gamma(1, r)$ and $\mathscr{L}(\eta_n) \to \Gamma(1, s)$.

1.34. The Jacobian of the transformation $y_1 = nx_1, y_2 = (n-1)(x_2 - x_1), \ldots, y_n = x_n - x_{n-1}$ is $J(x_1, \ldots, x_n) = n!$. Whence, by using (1.2) and taking into account the hint, we find that the joint distribution density of $Y_1, \ldots, Y_n$ is $\exp\{-y_1 - \ldots - y_n\}$. Since $X_{(k)} = \sum_{j=1}^{n} Y_j/(n-j+1)$, we have

$$\mathbf{E}X_{(k)} = \sum_{j=1}^{k} \frac{1}{n-j+1}\mathbf{E}Y_j, \quad \mathbf{D}X_{(k)} = \sum_{j=1}^{k} \frac{1}{(n-j+1)^2}\mathbf{D}Y_j.$$

The mean and variance of the exponential distribution $\Gamma(1, 1)$ are equal to 1, and we finally obtain

$$\mathbf{E}X_{(k)} = \sum_{j=n-k+1}^{n} \frac{1}{j}, \quad \mathbf{D}X_{(k)} = \sum_{j=n-k+1}^{n} \frac{1}{j^2}.$$

In particular, if $n \to \infty$, we have

$$\mathbf{E}X_{(n)} = \sum_{j=1}^{n} \frac{1}{j} = \ln n + c + o(1),$$

where $c = 0.5772\ldots$ is *Euler's constant*, and

$$\mathbf{D}X_{(n)} = \sum_{j=1}^{n} \frac{1}{j^2} = \frac{\pi^2}{6} + o(1).$$

1.35. The result of Problem 1.31 implies that the joint distribution density of the random variables $X_{(k)}$ and $X_{(l)}$ is

$$g_{kl}(x_1, x_2) = \frac{n!}{(k-1)!(l-k-1)!(n-l)!} x_1^{k-1}(x_2 - x_1)^{l-k-1}(1 - x_2)^{n-l},$$
$$0 \leqslant x_1 \leqslant x_2 \leqslant 1.$$

Then (see (1.2)), the joint distribution density for $Y_1 = X_{(k)}$ and $Y_2 = X_{(l)} - X_{(k)}$ has the form

$$\varphi(y_1, y_2) = \frac{n!}{(k-1)!(l-k-1)!(n-l)!} y_1^{k-1}y_2^{l-k-1}(1 - y_1 - y_2)^{n-l},$$
$$y_1, y_2 \geqslant 0, \ y_1 + y_2 \leqslant 1.$$

In order to find the distribution density for Y_2 it is sufficient to evaluate the integral

$$\int_0^{1-y_2} \varphi(y_1, y_2)\,dy_1 = \frac{n!}{(l-k-1)!(n-l+k)!} y_2^{l-k-1}(1-y_2)^{n-l+k},$$

$$0 \leqslant y_2 \leqslant 1.$$

Similarly, the distribution density for $X_{(k)}$ is

$$\int_0^{1-y_1} \varphi(y_1, y_2)\,dy_2 = \frac{n!}{(k-1)!(n-k)!} y_1^{k-1}(1-y_1)^{n-k},$$

$$0 \leqslant y_1 \leqslant 1.$$

Since the mean and variance of the distribution $B(a, b)$ are $\frac{a}{a+b}$ and $\frac{ab}{(a+b)^2(a+b+1)}$, respectively, we have

$$\mathbf{E}X_{(k)} = \frac{k}{n+1}, \quad \mathbf{D}X_{(k)} = \frac{k(n-k+1)}{(n+1)^2(n+2)},$$

$$\mathbf{E}(X_{(l)} - X_{(k)}) = \frac{l-k}{n+1}, \quad \mathbf{D}(X_{(l)} - X_{(k)}) = \frac{(l-k)(n-l+k+1)}{(n+1)^2(n+2)}.$$

Finally, since

$$\mathbf{D}(X_{(l)} - X_{(k)}) = \mathbf{D}X_{(k)} + \mathbf{D}X_{(l)} - 2\operatorname{cov}(X_{(k)}, X_{(l)}),$$

we find that

$$\operatorname{cov}(X_{(k)}, X_{(l)}) = \frac{k(n-l+1)}{(n+1)^2(n+2)}.$$

1.36. Noting that $\mathscr{L}\left(\frac{\xi - a}{b - a}\right) = R(0, 1)$, we may use the solution to the previous problem. In our case $X_{(1)} = (b-a)X'_{(1)} + a$, $X_{(n)} = (b-a)X'_{(n)} + a$, where $X'_{(1)}$ and $X'_{(n)}$ are the extrema of the sample of size n from the distribution $R(0, 1)$, whose joint density is

$$g_{1n}(x_1, x_2) = n(n-1)(x_2 - x_1)^{n-2}, \quad 0 \leqslant x_1 \leqslant x_2 \leqslant 1.$$

1.37. $\mathbf{P}(X_{(1)} > x) = \mathbf{P}(X_i > x,\ i = 1, \ldots, n) = [1 - F(x)]^n = e^{-n\left(\frac{x-a}{b}\right)^{\alpha}}$, $x \geqslant a$. From this we have

$$\mathbf{P}(X_{(1)} \leqslant x) = 1 - e^{-n\left(\frac{x-a}{b}\right)^{\alpha}}, \quad x \geqslant a,$$

and

$$\mathbf{P}(n^{1/\alpha}(X_{(1)} - a)/b \leqslant t) = 1 - e^{-t^{\alpha}}, \quad t \geqslant 0.$$

Thus, the random variable $n^{1/\alpha}(X_{(1)} - a)/b$ has the distribution $W(0, \alpha, 1)$ which is independent of the size of the sample. We infer that

$$\mathbf{E}X_{(1)} = a + b\Gamma\left(1 + \frac{1}{\alpha}\right)n^{-1/\alpha},$$

$$\mathbf{D}X_{(1)} = b^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]n^{-2/\alpha}.$$

1.38. The terms in $F_n(x_1, x_2)$ are independent and distributed like the random variable $\eta = I(\xi_1 \leqslant x_1)I(\xi_2 \leqslant x_2)$. Therefore,

$$\mathbf{E}F_n(x_1, x_2) = \mathbf{E}\eta = \mathbf{P}(\eta = 1) = \mathbf{P}(\xi_1 \leqslant x_1, \xi_2 \leqslant x_2) = F(x_1, x_2),$$

$$\mathbf{D}F_n(x_1, x_2) = \frac{1}{n}\mathbf{D}\eta = \frac{1}{n}[\mathbf{E}\eta - (\mathbf{E}\eta)^2] = \frac{1}{n}F(x_1, x_2)(1 - F(x_1, x_2)).$$

By Chebyshev's inequality we have

$$\mathbf{P}(|F_n(x_1, x_2) - F(x_1, x_2)| > \varepsilon) \leqslant \frac{1}{\varepsilon^2}\mathbf{D}F_n(x_1, x_2) \to 0$$

as $n \to \infty$. We use $\mathbf{X}_j = (X_{1j}, \ldots, X_{nj})$, $j = 1, 2$, $\overline{\mathbf{X}}_j$, $S_j^2 = S^2(\mathbf{X}_j)$ to denote the respective sample means and variances, and $S_{12} = \frac{1}{n}\sum_{i=1}^{n}(X_{i1} - \overline{X}_1) \times (X_{i2} - \overline{X}_2) = \frac{1}{n}\sum_{i=1}^{n}X_{i1}X_{i2} - \overline{X}_1\overline{X}_2$ to denote the sample covariance. Then the statistical analogue for the correlation coefficient $\varrho = \operatorname{cov}(\xi_1, \xi_2)/\sqrt{\mathbf{D}\xi_1\mathbf{D}\xi_2}$ will be $\varrho_n = S_{12}/S_1S_2$. If $\mathbf{E}(\xi_1^2\xi_2^2) < \infty$, then there exists $\mathbf{D}\left(\frac{1}{n}\sum_{i=1}^{n}X_{i1}X_{i2}\right) = \frac{1}{n}\mathbf{D}(\xi_1\xi_2)$ and it follows from Chebyshev's inequality that

$$\frac{1}{n}\sum_{i=1}^{n}X_{i1}X_{i2} \xrightarrow{\mathbf{P}} \mathbf{E}(\xi_1\xi_2)$$

for $n \to \infty$. Since $\overline{X_j} \xrightarrow{\mathbf{P}} \mathbf{E}\xi_j$, $S^2(\mathbf{X}_j) \xrightarrow{\mathbf{P}} \mathbf{D}\xi_j$, $j = 1, 2$, we have $\varrho_n \xrightarrow{\mathbf{P}} \varrho$ if $\mathbf{D}\xi_j > 0$, $j = 1, 2$.

1.39. (1) If $\mathcal{L}(\xi) = \mathcal{N}(\mu, \sigma^2)$, then $\mathbf{E}e^{it\xi} = e^{it\mu - \frac{\sigma^2}{2}t^2}$.

(2) If $\mathcal{L}(\xi) = \Gamma(a, \lambda)$, then $\mathbf{E}e^{it\xi} = \dfrac{1}{(1 - iat)^{\lambda}}$.

(3) If $\mathcal{L}(\nu_1, \ldots, \nu_N) = M(n;\ p_1, \ldots, p_N)$, then $\mathbf{E}(x_1^{\nu_1} \ldots x_N^{\nu_N}) = (x_1 p_1 + \ldots + x_N p_N)^n$.

(4) If $\mathcal{L}(\xi) = \Pi(\lambda)$, then $\mathbf{E}x^{\xi} = e^{\lambda(x - 1)}$.

(5) If $\mathcal{L}(\xi) = \overline{Bi}(r, p)$, then $\mathbf{E}x^{\xi} = \dfrac{q^r}{(1 - px)^r}$.

1.40. Suppose that $\mathbf{U}$ is an orthogonal matrix which reduces $\mathbf{\Sigma}$ to a diagonal form, i.e., $\mathbf{U}'\mathbf{\Sigma}\mathbf{U} = \mathbf{D}$. We write $\mathbf{B} = \mathbf{U}\mathbf{D}^{1/2}$. Then $\mathbf{\Sigma} = \mathbf{B}\mathbf{B}'$, and $\mathbf{Y} = \mathbf{B}\mathbf{X} + \boldsymbol{\mu}$, where the components of the vector $\mathbf{X}$ are independent and normal as $\mathcal{N}(0, 1)$. We have $Q = \mathbf{X}'\mathbf{B}'\mathbf{A}\mathbf{B}\mathbf{X} \equiv \mathbf{X}'\mathbf{A}_1\mathbf{X}$ and, as stated, $\mathbf{A}_1^2 = \mathbf{B}'\mathbf{A}\mathbf{B}\mathbf{B}'\mathbf{A}\mathbf{B} = \mathbf{B}'\mathbf{A}\mathbf{B} = \mathbf{A}_1$, i.e., the matrix $\mathbf{A}_1$ is idempotent. Consequently (see assertion 2° from Sec. 1.6), $\mathcal{L}(Q) = \chi^2(\operatorname{tr} \mathbf{A}_1)$. But $\operatorname{tr} \mathbf{A}_1 = \operatorname{tr} (\mathbf{A}\mathbf{B}\mathbf{B}') = \operatorname{tr} (\mathbf{A}\mathbf{\Sigma}) = m$, q.e.d.

1.44. The joint distribution density of X_1 and X_2 is

$$f(x_1, x_2) = \frac{x_1^{\lambda_1 - 1} x_2^{\lambda_2 - 1}}{a^{\lambda_1 + \lambda_2}\Gamma(\lambda_1)\Gamma(\lambda_2)} e^{-(x_1 + x_2)/a}, \quad x_1, x_2 > 0.$$

Consider the transformation $y_1 = x_1 + x_2$, $y_2 = x_1/(x_1 + x_2)$. It uniquely maps the region $\{x_1, x_2 > 0\}$ onto the region $\{y_1 > 0,\ 0 < y_2 < 1\}$, and its Jacobian is $J(x_1, x_2) = -1/(x_1 + x_2)$. By formula (1.2) the joint distribution density of Y_1 and Y_2 is

$$\varphi(y_1, y_2) = f(y_1 y_2,\ y_1(1 - y_2))y_1 = \frac{y_1^{\lambda_1 + \lambda_2 - 1} e^{-y_1/a}}{\Gamma(\lambda_1 + \lambda_2)a^{\lambda_1 + \lambda_2}} \frac{y_2^{\lambda_1 - 1}(1 - y_2)^{\lambda_2 - 1}}{B(\lambda_1, \lambda_2)}.$$

1.45. The formula for the moments follows from that for the moments of the distribution $\Gamma(a, \lambda)$ for $a = 2$, $\lambda = n/2$, and the property $\Gamma(\lambda + 1) = \lambda\Gamma(\lambda)$. Specifically, $\mathbf{E}\chi_1^2 = 1$, $\mathbf{D}\chi_1^2 = 2$. By the reproducibility property of the gamma distribution we have $\chi_n^2 = \xi_1 + \ldots + \xi_n$, where the terms are independent and equally distributed as $\Gamma(2, 1/2) = \chi^2(1)$. By the Central Limit Theorem as $n \to \infty$ the random variable $(\chi_n^2 - n)/\sqrt{2n}$ is asymptotically normal as $\mathcal{N}(0, 1)$.

1.46. The first assertion directly follows from (1.4). In the second case we have

$$\mathbf{P}(a + \tan \xi \leqslant x) = \mathbf{P}(\xi \leqslant \arctan (x - a)) = \frac{1}{\pi}\left(\frac{\pi}{2} + \arctan (x - a)\right).$$

The sought-for distribution density is

$$\frac{1}{\pi} \arctan' (x - a) = \frac{1}{\pi} \frac{1}{1 + (x - a)^2}.$$

1.47. Since $\mathbf{E}t_n^{2r} = n^r\mathbf{E}\eta^{2r}\mathbf{E}(\chi_n^2)^{-r}$, the general formulas for the moments of the distributions $\mathcal{N}(0, 1)$ and $\Gamma(2, n/2)$ give

$$\mathbf{E}\eta^{2r} = 1 \times 3 \times \ldots \times (2r - 1),$$

$$\mathbf{E}(\chi_n^2)^{-r} = \frac{\Gamma\left(\frac{n}{2} - r\right)}{2^r\Gamma\left(\frac{n}{2}\right)} = \frac{1}{(n-2)(n-4)\ldots(n-2r)}$$

for $2r < n$. The remaining assertions about the moments follow from the form of the density $s_n(x)$. The assertion about the convergence of $s_n(x)$ follows from the relations

$$n^{-1/2}\Gamma\left(\frac{n+1}{2}\right)\Big/\Gamma\left(\frac{n}{2}\right) \to 1/\sqrt{2}, \qquad \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \to e^{-x^2/2}.$$

Finally, by the law of large numbers we have $\chi_n^2/n \xrightarrow{\mathbf{P}} 1$. Then $\sqrt{n/\chi_n^2} \xrightarrow{\mathbf{P}} 1$ and, consequently, $\mathcal{L}(t_n) \to \mathcal{L}(\eta) = \mathcal{N}(0, 1)$ (see the assertions 1° and 2° (c) from Sec. 1.4).

1.48. Take $Y = \chi_{n_1}^2/(\chi_{n_1}^2 + \chi_{n_2}^2)$. Then $F_{n_1,n_2} = \dfrac{n_2}{n_1}\dfrac{\chi_{n_1}^2}{\chi_{n_2}^2} = \dfrac{n_2}{n_1}\dfrac{Y}{1 - Y}$.

But (see Problem 1.44) $\mathcal{L}(Y) = B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$ and, therefore,

$$\mathbf{P}(F_{n_1,n_2} \leqslant x) = \mathbf{P}\left(Y \leqslant \frac{n_1x}{n_2 + n_1x}\right) = \mathrm{B}\left(\frac{n_1x}{n_2 + n_1x}; \frac{n_1}{2}, \frac{n_2}{2}\right).$$

Since $Y = \dfrac{F_{n_1,n_2}}{F_{n_1,n_2} + \frac{n_2}{n_1}}$, we obtain the relation $\mathcal{L}\left(\dfrac{F_{n_1,n_2}}{F_{n_1,n_2} + \frac{n_2}{n_1}}\right) = B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$. The moments can be calculated from $\mathbf{E}F_{n_1,n_2}^r = \left(\frac{n_2}{n_1}\right)^r\mathbf{E}(\chi_{n_1}^2)^r\mathbf{E}(\chi_{n_2}^2)^{-r}$ using the formula for the moments of the gamma distribution. The moments only exist for $-\frac{n_1}{2} < r < \frac{n_2}{2}$ and are equal to

$$\mathbf{E}F_{n_1,n_2}^r = \left(\frac{n_2}{n_1}\right)^r \frac{\Gamma\left(\frac{n_1}{2} + r\right)}{\Gamma\left(\frac{n_1}{2}\right)} \frac{\Gamma\left(\frac{n_2}{2} - r\right)}{\Gamma\left(\frac{n_2}{2}\right)}.$$

Specifically, for $n_2 > 2$ we have $\mathbf{E}F_{n_1,n_2} = \dfrac{n_2}{n_2 - 2}$, and for $n_2 > 4$ we have $\mathbf{D}F_{n_1,n_2} = \dfrac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$.

1.49. By the theorem about the mean we have

$$1 - \mathrm{B}(x; a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} c^{a-1} \frac{(1 - x)^b}{b}, \ c \in [x, 1].$$

For $b \to \infty$ and fixed a, Stirling's formula gives $\dfrac{\Gamma(a + b)}{\Gamma(b)} \sim b^a$. Consequently,

$$\frac{1}{b} \ln [1 - \mathrm{B}(x; a, b)] = \ln (1 - x) - \frac{1}{b} \ln b$$
$$+ \frac{1}{b} \ln \frac{\Gamma(a + b)}{\Gamma(b)} + \frac{1}{b} \ln \frac{c^{a-1}}{\Gamma(a)} \to \ln (1 - x).$$

The second relation follows from the first one and Problem 1.48.

1.50. The distribution of the random variable $t_n = \eta/\sqrt{\chi_n^2/n}$ is symmetric (because the distributions of $-\eta$ and η coincide), therefore,

$$\mathbf{P}(t_n^2 \leqslant x^2) = \mathbf{P}(-|x| \leqslant t_n \leqslant |x|) = 2\mathbf{P}(t_n \leqslant |x|) - 1$$

or $\mathbf{P}(t_n \leqslant |x|) = \dfrac{1}{2} + \dfrac{1}{2}\mathbf{P}(t_n^2 \leqslant x^2)$. For $x > 0$ we get $\mathbf{P}'(t_n \leqslant x) = \dfrac{1}{2}\mathbf{P}'(t_n^2 \leqslant x^2)$, i.e., $s_n(x) = xf_{1,n}(x^2)$. These relations also imply that $\mathbf{P}(t_n > d\sqrt{n}) = [1 - F(d^2n; 1, n)]/2$. This and Problem 1.49 give the required limiting relation.

1.51. Since $\mathscr{L}\left(\dfrac{2}{a}(X_1 + \ldots + X_l)\right) = \chi^2(2l)$, $\mathscr{L}\left(\dfrac{2}{a}(X_{l+1} + \ldots + X_{l+m})\right) = \chi^2(2m)$, and the random variables are independent, the required assertion follows from the definition of Snedecor's law.

1.52. We write $\varphi(x_1, \ldots, x_N) = \mathbf{E}(x_1^{\nu_1} \ldots x_N^{\nu_N}) = (x_1p_1 + \ldots + x_Np_N)^n$. Then

$$\mathbf{E}(x_1^{\nu_1} \ldots x_k^{\nu_k}) = \varphi(x_1, \ldots, x_k, 1, \ldots, 1)$$
$$= (x_1p_1 + \ldots + x_kp_k + 1 - p_1 - \ldots - p_k)^n.$$

We have

$$\mathbf{E}(\nu_1)_{k_1} \ldots (\nu_N)_{k_N} = \frac{\partial^{k_1 + \ldots + k_N}}{\partial x_1^{k_1} \ldots \partial x_N^{k_N}} \varphi(x_1, \ldots, x_N)|_{x_1 = \ldots = x_N = 1}.$$

The direct calculation of the derivative leads to the sought-for formula. Finally,

$$\mathbf{E}\eta^j = \sum_{l=1}^{N} c_l^j np_l = n\overline{c^j},$$

$$\operatorname{cov}(\eta^1, \eta^2) = \sum_{i,j=1}^{N} c_i^1 c_j^2 \operatorname{cov}(\nu_i, \nu_j) = n \sum_{i=1}^{N} c_i^1 c_i^2 p_i(1 - p_i)$$

$$-n \sum_{i \neq j} c_i^1 c_j^2 p_i p_j = n\left(\sum_{i=1}^{N} c_i^1 c_i^2 p_i - \sum_{i,j=1}^{N} c_i^1 c_j^2 p_i p_j \right) = n(\overline{c^1 c^2} - \overline{c}^1 \overline{c}^2).$$

1.53. We write $\boldsymbol{\nu}^* = (\nu_1^*, \ldots, \nu_k^*)$, where $\nu_j^* = (\nu_j - np_j)/\sqrt{n}$, $j = 1, \ldots, k$. It is sufficient to show that the characteristic function is

$$\mathbf{E}e^{i\mathbf{t}'\boldsymbol{\nu}^*} \to \exp\left\{-\frac{1}{2}\mathbf{t}'\Sigma_k\mathbf{t}\right\}, \quad \mathbf{t} = (t_1, \ldots, t_k),$$

for any fixed $\mathbf{t}$. It follows from the previous problem that

$$\mathbf{E}e^{i\mathbf{t}'\boldsymbol{\nu}^*} = e^{-i\sqrt{n}\mathbf{t}'\mathbf{p}}\left[1 + \sum_{j=1}^{k} p_j(e^{it_j/\sqrt{n}} - 1)\right]^n, \quad \mathbf{p} = (p_1, \ldots, p_k).$$

We take the logarithm of this expression and recall that $\ln(1 + \varepsilon) = \varepsilon - \frac{\varepsilon^2}{2} + O(\varepsilon^3)$, $\varepsilon \to 0$. Under the conditions of the problem we get

$$\ln \mathbf{E}e^{i\mathbf{t}'\boldsymbol{\nu}^*} = -i\sqrt{n}\mathbf{t}'\mathbf{p} + n\sum_{j=1}^{k} p_j(e^{it_j/\sqrt{n}} - 1)$$

$$-\frac{n}{2}\left[\sum_{j=1}^{k} p_j(e^{it_j/\sqrt{n}} - 1)\right]^2 + O\left(\frac{1}{\sqrt{n}}\right)$$

$$= -\frac{1}{2}\sum_{j=1}^{k} p_j t_j^2 + \frac{1}{2}\left(\sum_{j=1}^{k} p_j t_j\right)^2 + O\left(\frac{1}{\sqrt{n}}\right)$$

$$= -\frac{1}{2}\mathbf{t}'\Sigma_k\mathbf{t} + O\left(\frac{1}{\sqrt{n}}\right),$$

q.e.d.

Finally, we get after some algebra

$$|\Sigma_k| = p_1 \ldots p_k(1 - p_1 - \ldots - p_k) \neq 0 \quad \text{for} \quad k < N.$$

1.54. For any integer non-negative $k_1, \ldots, k_N$, such that $k_1 + \ldots + k_N = n$, we get

$$\mathbf{P}(\xi_j = k_j, j = 1, \ldots, N | \xi_1 + \ldots + \xi_N = n) = \frac{\mathbf{P}(\xi_j = k_j, j = 1, \ldots, N)}{\mathbf{P}(\xi_1 + \ldots + \xi_N = n)}.$$

Since $\xi_1, \ldots, \xi_N$ are independent, the numerator is equal to

$$\prod_{j=1}^{N} e^{-\lambda_j}\lambda_j^{k_j}/k_j! = e^{-\lambda}\prod_{j=1}^{N} \frac{\lambda_j^{k_j}}{k_j!}, \quad \lambda = \lambda_1 + \ldots + \lambda_N.$$

We know (see Problem 1.39 (4)) that $\mathscr{L}(\xi_1 + \ldots + \xi_N) = \Pi(\lambda)$, and the denominator is $e^{-\lambda}\lambda^n/n!$. The sought-for probability is $\frac{n!}{k_1!\ldots k_N!} p_1^{k_1}\ldots p_N^{k_N}$, which proves the assertion.

1.55. Compute the unconditional probabilities $\mathbf{P}(\xi = k)$, $k = 0, 1, \ldots$. We have

$$\mathbf{P}(\xi = k) = \int_0^\infty e^{-\lambda}\frac{\lambda^k}{k!}\frac{\lambda^{r-1}}{\Gamma(r)a^r}e^{-\lambda/a}\,d\lambda$$

$$= \frac{\Gamma(k+r)}{k!\Gamma(r)}\left(\frac{a}{a+1}\right)^{k+r}\frac{1}{a^r} = C_{r+k-1}^k\left(\frac{a}{a+1}\right)^k\frac{1}{(a+1)^r}.$$

1.56. The vector $(\overline{X}, X_1 - \overline{X}, \ldots, X_n - \overline{X})$ is distributed normally because it is a linear transformation of the normal vector $\mathbf{X}$. Here $\operatorname{cov}(\overline{X}, X_i - \overline{X}) = \operatorname{cov}(\overline{X}, X_i) - \mathbf{D}\overline{X} = \frac{1}{n}\mathbf{D}X_i - \mathbf{D}\overline{X} = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$, $i = 1, \ldots, n$. Therefore, the first component is independent of all the others.

1.57. Suppose that $\mathbf{U}$ is an orthogonal matrix which reduces $\mathbf{A}_1$ to a diagonal form, i.e., $\mathbf{U}'\mathbf{A}_1\mathbf{U} = \mathbf{D}_1$, where, as stated, $n - n_1 = n_2$ diagonal elements of $\mathbf{D}_1$ are zero. We introduce the vector $\eta = \mathbf{U}'\mathbf{X}$. Then $\mathbf{X} = \mathbf{U}\eta$ and we may write

$$Q = \eta\eta' = \eta'\mathbf{U}'\mathbf{A}_1\mathbf{U}\eta + \eta'\mathbf{U}'\mathbf{A}_2\mathbf{U}\eta = \eta'\mathbf{D}_1\eta + \eta'\mathbf{D}_2\eta,$$

where $\mathbf{D}_2 = \mathbf{U}'\mathbf{A}_2\mathbf{U} = \mathbf{E}_n - \mathbf{D}_1$ is a diagonal matrix and rank $\mathbf{D}_2 =$ rank $\mathbf{A}_2 = n_2$. It follows that the diagonal elements of $\mathbf{D}_2$ which correspond to the zero elements of $\mathbf{D}_1$ are equal to unity and all the other elements are zero. This means that all the non-zero elements of $\mathbf{D}_1$ are equal to unity. Consequently, the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are idempotent. It follows from the above that $\mathbf{D}_1\mathbf{D}_2 = 0$ and hence $\mathbf{A}_1\mathbf{A}_2 = 0$.

1.58. If we go to the normalized quantities $X_i' = \frac{X_i - \mu}{\sigma}$, the form of η does not change, and we may consider that $(\mu, \sigma) = (0, 1)$. Let $\mathbf{B}$ be an $n \times n$ matrix whose all elements are $1/n$. Then $nS^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 = \mathbf{X}'\mathbf{A}\mathbf{X}$, where the matrix $\mathbf{A} = \mathbf{E}_n - \mathbf{B}$ is idempotent, and hence rank $\mathbf{A} =$ tr $\mathbf{A} = n - 1$. It follows that $n - 1$ of the eigenvalues of $\mathbf{A}$ are unities and

one eigenvalue is zero. Suppose that $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ are the eigenvectors of $\mathbf{A}$, which correspond to the eigenvalue 1. Then $\mathbf{A} = \sum_{k=1}^{n-1} \mathbf{u}_k\mathbf{u}_k'$ is a spectral representation of $\mathbf{A}$, and we can directly check that $\mathbf{u}_1 = \sqrt{\frac{n}{n-1}}\left(\frac{n-1}{n}, -\frac{1}{n}, \ldots, -\frac{1}{n}\right)$, $\sqrt{\frac{n}{n-1}}(X_1 - \overline{X}) = \mathbf{u}_1'\mathbf{X}$, and $nS^2 = \sum_{k=1}^{n-1}(\mathbf{u}_k'\mathbf{X})^2$. Thus, by writing $Y_k = \mathbf{u}_k'\mathbf{X}$, $k = 1, \ldots, n-1$, we obtain

$$\eta = \frac{X_1 - \overline{X}}{\sqrt{n-1}S} = \frac{Y_1}{\sqrt{Y_1^2 + \ldots + Y_{n-1}^2}},$$

where $Y_1, \ldots, Y_{n-1}$ are independent and $\mathcal{N}(0, 1)$-normal, which can be written as $\eta = Y_1/\sqrt{Y_1^2 + \chi_{n-2}^2}$. Here χ_{n-2}^2 is independent of Y_1 and $\mathcal{L}(\chi_{n-2}^2) = \chi^2(n-2)$. Now, by direct calculations, we can find the η-distribution. Note that this distribution is in the interval $(-1, 1)$ (since $\eta^2 < 1$) and is symmetric (since the distributions of $-Y_1$ and Y_1 coincide). For $0 < u < 1$ we will therefore have

$$\mathbf{P}(\eta > u) = \frac{1}{2}\mathbf{P}(\eta^2 > u^2) = \frac{1}{2}\mathbf{P}\left(\frac{\chi_{n-2}^2}{(n-2)Y_1^2} < \frac{1-u^2}{(n-2)u^2}\right)$$
$$= \frac{1}{2}F\left(\frac{1-u^2}{(n-2)u^2}; n-2, 1\right),$$

where $F(x; n_1, n_2)$ is the distribution function of Snedecor's law $S(n_1, n_2)$. Using the result of Problem 1.48, we write

$$F\left(\frac{1-u^2}{(n-2)u^2}; n-2, 1\right) = \mathrm{B}\left(1-u^2; \frac{n-2}{2}, \frac{1}{2}\right).$$

We finally find for $0 < u < 1$ that

$$F_\eta(u) = 1 - \frac{1}{2}\mathrm{B}\left(1-u^2; \frac{n-2}{2}, \frac{1}{2}\right).$$

For the negative values of u we have $F_\eta(u) = 1 - F_\eta(-u)$.

1.59. (a) The population of random variables $(\overline{X}_1, \overline{X}_2, X_{i1} - \overline{X}_1, X_{i2} - \overline{X}_2, i = 1, \ldots, n)$ is distributed normally because it is a linear transformation of the normal vector $(X_{i1}, X_{i2}, i = 1, \ldots, n)$. We can directly check that the first two components $\overline{X}_1$ and $\overline{X}_2$ are uncorrelated with the others. This means that $(\overline{X}_1, \overline{X}_2)$ and $(X_{i1} - \overline{X}_1, X_{i2} - \overline{X}_2, i = 1, \ldots, n)$ are independent and hence $(\overline{X}_1, \overline{X}_2)$ and (S_1^2, S_{12}, S_2^2) are also independent.

(b) It follows from the above that $\mathcal{L}(\overline{X}_1, \overline{X}_2) = \mathcal{N}\left((\mu_1, \mu_2), \frac{1}{n}\Sigma\right)$. By assuming in Problem 1.40 that $\mathbf{A} = n\Sigma^{-1}$, we will obtain the required result.

(c) We have $S_1^2 = \frac{\sigma_1^2}{n}(Y_1^2 + Y_2)$, $S_{12} = \frac{\sigma_1\sigma_2}{n} Y_1\sqrt{Y_3}$, $S_2^2 = \frac{\sigma_2^2}{n} Y_3$, and the absolute value of the transformation's Jacobian is

$$|J| = \frac{n^{5/2}}{\sigma_1^3\sigma_2^2\sqrt{x_2}} = \frac{n^3}{(\sigma_1\sigma_2)^3\sqrt{y_3}}.$$

By (1.2) the density of the joint distribution of (Y_1, Y_2, Y_3) is

$$\varphi(y_1, y_2, y_3) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2}{2}} y_2^{\frac{n-4}{2}} e^{-\frac{y_2}{2}} y_3^{\frac{n-3}{2}} e^{-\frac{y_3}{2}} / 2\sqrt{2\pi}\Gamma(n-2),$$

i.e., it is the product of the densities of the distributions $\mathcal{N}(0, 1)$, $\chi^2(n-2)$, and $\chi^2(n-1)$ $\Big($the coefficient is reduced to the required form by the formula $\Gamma(p)\Gamma\left(p + \frac{1}{2}\right)2^{2p-1} = \sqrt{\pi}\Gamma(2p)\Big)$. Thus, the random variables Y_1, Y_2, and Y_3 are independent and $\mathcal{L}(Y_1) = \mathcal{N}(0, 1)$, $\mathcal{L}(Y_2) = \chi^2(n-2)$, $\mathcal{L}(Y_3) = \chi^2(n-1)$. Whence

$$\mathcal{L}(T = Y_1/\sqrt{Y_2/(n-2)}) = S(n-2).$$

Since $\varrho_n = \frac{T}{\sqrt{n-2+T^2}}$, by using (1.3), we finally find that the distribution density of the random variable ϱ_n is

$$\varphi(y) = s_{n-2}\left(\frac{\sqrt{n-2}\,y}{\sqrt{1-y^2}}\right)\frac{\sqrt{n-2}}{(1-y^2)^{3/2}} = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)}(1-y^2)^{(n-4)/2},$$

$$-1 < y < 1.$$

1.60. It follows from the solution of Problem 1.29 that the asymptotic distributions of the statistics S^2 and A_{n2} are similar, and hence (see Problem 1.28) the joint distribution of $\overline{X}$ and S^2 is asymptotically normal. The distribution of any of their linear combination is also asymptotically normal, and therefore it is sufficient to compute the mean and variance of the difference $\overline{X} - \frac{n}{n-1}S^2$. Using the solution of Problem 1.27, we find

$$\mathbf{E}\left(\overline{X} - \frac{n}{n-1}S^2\right) = 0,$$

$$\mathbf{D}\left(\overline{X} - \frac{n}{n-1}S^2\right) = \mathbf{D}\overline{X} + \left(\frac{n}{n-1}\right)^2 \mathbf{D}S^2 - \frac{2n}{n-1}\operatorname{cov}(\overline{X}, S^2)$$

$$= \frac{\mu_2}{n} + \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right) - \frac{\mu_3}{n} = \frac{2\lambda^2}{n-1},$$

whence we conclude that $\mathscr{L}(\zeta_n) \to \mathscr{N}(0, 1)$. But $T_n = \zeta_n \overline{X}/\lambda$ and, taking into account the hint, we arrive at the required result.

1.62. Here $\mathscr{L}(\xi_1) = \mathscr{N}(\mu_1, \sigma_1^2)$ and $f_{\xi_1}(x) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}$. We have

$$f_{\xi_1\xi_2}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}}$$

$$\times \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\varrho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$

Simple calculations lead to the required expression for the conditional density $f_{\xi_2|\xi_1}(y|x)$.

1.63. Let $\mathscr{L} = \left\|\begin{matrix} \mathbf{E}^{(1)} & \mathbf{0} \\ -\mathbf{A} & \mathbf{E}^{(2)} \end{matrix}\right\|$ be the matrix of the transformation. Then

$\mathscr{L}(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}) = \mathscr{N}\left(\mathbf{L}\begin{pmatrix}\mu^{(1)} \\ \mu^{(2)}\end{pmatrix}, \mathbf{L}\Sigma\mathbf{L}'\right)$ and direct calculations give

$$\mathbf{L}\Sigma\mathbf{L}' = \left\|\begin{matrix} \mathbf{E}^{(1)} & \mathbf{0} \\ -\mathbf{A} & \mathbf{E}^{(2)} \end{matrix}\right\| \times \left\|\begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix}\right\| \times \left\|\begin{matrix} \mathbf{E}^{(1)} & -\mathbf{A}' \\ \mathbf{0} & \mathbf{E}^{(2)} \end{matrix}\right\| = \left\|\begin{matrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{matrix}\right\|$$

(in particular, this leads to the equation $|\Sigma| = |\Sigma_{11}| \times |\mathbf{B}|$). The structure of the matrix $\mathbf{L}\Sigma\mathbf{L}'$ implies that $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are uncorrelated and, consequently, independent. Besides we find that $\mathscr{L}(\mathbf{Y}^{(2)}) = \mathscr{N}(\mu_2 - \mathbf{A}\mu^{(1)}, \mathbf{B})$. Then

$$\mathscr{L}(\mathbf{X}^{(2)}|\mathbf{X}^{(1)} = \mathbf{x}^{(1)}) = \mathscr{L}(\mathbf{Y}^{(2)} + \mathbf{A}\mathbf{x}^{(1)}) = \mathscr{N}(M(\mathbf{x}^{(1)}), \mathbf{B}).$$

1.64. By the formula for the total probability, the probability of the event is

$$\mathbf{P}(\xi = k) = \int_0^\lambda \mathbf{P}'\left(-\sum_{i=1}^k \ln U_i \leqslant t\right)\mathbf{P}(-\ln U_{k+1} > \lambda - t)\,dt$$

$$= \int_0^\lambda \frac{t^{k-1}e^{-t}}{(k-1)!}e^{-\lambda+t}\,dt = e^{-\lambda}\frac{\lambda^k}{k!}.$$

1.65. Since $\mathbf{P}(cU_1 \leqslant f(x)) = f(x)/c$ and

$$\mathbf{P}(cU_1 > f(a + (b - a)U_2)) = \int_0^1 \left(1 - \frac{1}{c} f(a + (b - a)x)\right) dx$$
$$= 1 - \frac{1}{c(b - a)},$$

the distribution density of the random variable ξ at the point $x \in [a, b]$ is calculated by the formula

$$f_\xi(x) = \frac{1}{c(b - a)} f(x) \sum_{r=0}^{\infty} \left(1 - \frac{1}{c(b - a)}\right)^r = f(x).$$

TO CHAPTER 2

2.1. The problems are solved by calculating the means and variances of the indicated statistics and require the asymptotic analysis of these characteristics as $n \to \infty$. For example, in problem (a) we have $F_n(x) = \mu_n(x)/n$, where $\mu_n(x)$ is the number of units in the sample $\mathbf{X} = (X_1, \ldots, X_n)$, which assumed the values $\leqslant x$, i.e., $\mathscr{L}(\mu_n(x)) = Bi(n, F(x))$, whence we have for $n \to \infty$

$$\mathbf{E}F_n(x) = \mathbf{E}\mu_n(x)/n = F(x),$$
$$\mathbf{D}F_n(x) = \mathbf{D}\mu_n(x)/n^2 = F(x)(1 - F(x))/n \to 0$$

which means that $F_n(x)$ is an unbiased and consistent estimator for $F(x)$.

Consider problem (d). Since (see Problem 1.27)

$$\mathbf{E}S^2 = \frac{n - 1}{n} \mu_2 = \mu_2 + O\left(\frac{1}{n}\right),$$

$$\mathbf{D}S^2 = \frac{(n - 1)^2}{n^3} \left(\mu_4 - \frac{n - 3}{n - 1} \mu_2^2\right) = O\left(\frac{1}{n}\right) \quad \text{for} \quad \mu_4 < \infty,$$

S^2 is a consistent and biased estimator for μ_2. To eliminate the bias, we must use the statistic $\dfrac{n}{n - 1} S^2 \equiv S'^2$ whose variance is $\left(\dfrac{n}{n - 1}\right)^2 \mathbf{D}S^2 = O\left(\dfrac{1}{n}\right)$, i.e., S'^2 is also a consistent estimator for μ_2.

2.2. The solution of Problem 2.1 (b) gives

$$T_n(\mathbf{X}) \xrightarrow{\mathbf{P}} \sqrt{\alpha_2/2} = \alpha_1 \Leftrightarrow \sqrt{\mu_2} = \alpha_1,$$

i.e., the standard deviation and the mean are equal. Specifically, this is true for the distributions $\Gamma(a, 1)$ and $\mathscr{N}(a, a^2)$, $a > 0$.

2.4. For the random variable $\eta = \xi_1 + \xi_2$ the sampling data are $(Y_i = X_{i1} + X_{i2},\ i = 1, \ldots, n)$, and its variance is $\mathbf{D}\eta = \mathbf{D}\xi_1 + \mathbf{D}\xi_2 + 2\operatorname{cov}(\xi_1, \xi_2)$. Problem 2.1 (d) implies that the statistic

$$\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{i1} - \overline{X}_1)^2$$

$$+ \frac{1}{n-1}\sum_{i=1}^{n}(X_{i2} - \overline{X}_2)^2 + \frac{2}{n-1}\sum_{i=1}^{n}(X_{i1} - \overline{X}_1)(X_{i2} - \overline{X}_2)$$

is an unbiased estimator for $\mathbf{D}\eta$. The first two sums of this expansion are unbiased estimators for the variances $\mathbf{D}\xi_1$ and $\mathbf{D}\xi_2$, respectively. Therefore,

$$\mathbf{E}\,\frac{1}{n-1}\sum_{i=1}^{n}(X_{i1} - \overline{X}_1)(X_{i2} - \overline{X}_2) = \operatorname{cov}(\xi_1, \xi_2),$$

q.e.d.

2.5. For an arbitrary statistic $T(\mathbf{X})$ we have

$$\mathbf{E}_\theta T(\mathbf{X}) = \sum_{\substack{\mathbf{x} = (x_1, \ldots, x_n)\\ x_i = 0, 1,\\ i = 1, \ldots, n}} T(\mathbf{x})\theta^{\Sigma x_i}(1-\theta)^{n - \Sigma x_i}$$

$$= \sum_{r=0}^{n} \theta^r(1-\theta)^{n-r} \sum_{\mathbf{x}:\ \Sigma x_i = r} T(\mathbf{x}).$$

The right-hand side of this expression is a polynomial in θ of degree $\leqslant n$. Consequently, the unbiased estimators in this model can only be constructed for the polynomials $\tau(\theta) = \sum_{k=0}^{s} a_k\theta^k$ for $s \leqslant n$.

2.6. Compute the first two moments of the statistic T. Since $\mathscr{L}_\theta(r_n) = Bi(n, \theta)$, we have $\mathbf{E}_\theta T = \dfrac{1}{n+\beta}(\mathbf{E}_\theta r_n + \alpha) = \dfrac{n\theta + \alpha}{n+\beta}$ and

$$\mathbf{E}_\theta T^2 = \frac{1}{(n+\beta)^2}(\mathbf{D}_\theta r_n + (\mathbf{E}_\theta r_n)^2 + 2\alpha\mathbf{E}_\theta r_n + \alpha^2)$$

$$= \frac{n(n-1)\theta^2 + (2\alpha+1)n\theta + \alpha^2}{(n+\beta)^2}.$$

We find

$$\Delta(\alpha, \beta; \theta) \equiv \mathbf{E}_\theta(T-\theta)^2 = \mathbf{E}_\theta T^2 - 2\theta\mathbf{E}_\theta T + \theta^2$$

$$= \frac{\theta^2(\beta^2 - n) + \theta(n - 2\alpha\beta) + \alpha^2}{(n+\beta)^2}.$$

Specifically, $\Delta(0, 0; \theta) = \mathbf{D}_\theta\left(\frac{r_n}{n}\right) = \frac{\theta(1-\theta)}{n}$, $\Delta\left(\frac{\sqrt{n}}{2}, \sqrt{n}; \theta\right) = \frac{1}{4(\sqrt{n}+1)^2}$ is independent of θ.

Consider the estimator $T' = \dfrac{r_n + \sqrt{n}/2}{n + \sqrt{n}}$. Its mean square error is smaller than that of the unbiased estimator $T^* = \dfrac{r_n}{n}$, i.e., $\dfrac{1}{4(\sqrt{n}+1)^2} < \dfrac{\theta(1-\theta)}{n}$ for $\theta \in \left(\dfrac{1}{2} \pm \dfrac{\sqrt{2\sqrt{n}+1}}{2(\sqrt{n}+1)}\right)$. The length of this interval tends to zero as $n \to \infty$. For the other values of the parameter $\theta \in (0, 1)$ the estimator T^* is more exact. Thus, the estimators T' and T^* cannot be compared by the mean-square-error test, and we must have other reasons to choose one of them. For example, we may consider that the estimator with a smaller maximum value of the mean square error is better (the *minimax principle*). Since $\max\limits_\theta \theta(1-\theta) = \dfrac{1}{4}$ and $\dfrac{1}{4(\sqrt{n}+1)^2} < \dfrac{1}{4n}$, the minimax principle implies that T' is better than T^*.

2.7. The unbiasedness condition $\mathbf{E}_\theta T(X) = \tau_{rs}(\theta)\ \forall\theta \in (0, 1)$ has here the form

$$\sum_{j=0}^{k} T(j) C_k^j \theta^j (1-\theta)^{k-j} = \theta^r (1-\theta)^s \quad \forall\theta \in (0, 1).$$

For any T the left-hand side of this identity is a polynomial in θ of degree $\leqslant k$. Consequently, the identity is only true for $r + s \leqslant k$. Now we can directly check whether the statistic in the statement of the problem meets the unbiasedness condition and then show that this unbiased estimator is unique. We suppose that $T'(X)$ is another unbiased estimator. Then the statistic $T_1(X) = T(X) - T'(X)$ satisfies the identity

$$\sum_{j=0}^{k} T_1(j) C_k^j \theta^j (1-\theta)^{k-j} \equiv 0, \quad 0 < \theta < 1,$$

or $\sum\limits_{j=0}^{k} T_1(j) C_k^j x^j \equiv 0$, $0 < x < \infty$, where $x = \dfrac{\theta}{1-\theta}$. But since the polynomial is identically zero, all its coefficients are zero, i.e., $T'(j) = T(j)$ for $j = 0, 1, \ldots, k$.

2.8. Due to the reproducibility of the binomial distribution we have $\mathscr{L}_\theta(T) = Bi(kn, \theta)$ and, therefore,

$$\mathbf{E}_\theta H(T) = \sum_{j=0}^{kn} H(j) C_{kn}^j \theta^j (1-\theta)^{kn-j}.$$

For any function H this mean is a polynomial in θ of degree $\leqslant kn$. Consequently, the unbiased estimators of the form $H(T)$ can only be constructed in this

model for the functions of the form $\tau(\theta) = \sum_{r=0}^{s} a_r\theta^r$ for $s \leqslant kn$. Suppose that $\tau_j(\theta) = \theta^j, j \leqslant kn$. Since $\mathbf{E}_\theta(T)_j = (kn)_j\theta^j$ (see Problem 1.52), the unbiased estimator for $\tau_j(\theta)$ is the statistic $\tau_j^* = (T)_j/(kn)_j$. Using the technique from the previous problem, we can show that this is a unique unbiased estimator which depends on T.

2.9. The first assertion follows from the chain of equations

$$\mathbf{E}_\theta(X)_j = \sum_{k=0}^{\infty}(k)_j e^{-\theta}\frac{\theta^k}{k!} = \theta^j\sum_{k=j}^{\infty} e^{-\theta}\frac{\theta^{k-j}}{(k-j)!} = \theta^j.$$

To prove the second assertion, we write the unbiasedness condition $\mathbf{E}_\theta T(X) = \tau(\theta)\ \forall\theta > 0$ in the form

$$\sum_{k=0}^{\infty} T(k)\frac{\theta^{k+a}}{k!} = \sum_{r=0}^{\infty}\frac{\theta^r}{r!} \quad \forall\theta > 0.$$

It is clear that there is no function $T(k)$ independent of θ, which satisfies this identity.

2.10. Here the unbiasedness condition $\mathbf{E}_\theta T(X) = \tau(\theta)\ \forall\theta > 0$ has the form

$$\sum_{k=0}^{\infty} T(k)\frac{\theta^k}{k!} = e^\theta(1 - e^{-\theta})^2 = e^\theta + e^{-\theta} - 2 = 2\sum_{r=1}^{\infty}\frac{\theta^{2r}}{(2r)!} \quad \forall\theta > 0.$$

The only function $T(k)$ which satisfies this identity has the form

$$T(k) = \begin{cases} 2 & \text{for even } k > 0, \\ 0 & \text{otherwise.} \end{cases}$$

This unbiased estimator for $T(X)$ is practically useless.

2.11. Here the unbiasedness condition

$$\sum_{k=0}^{\infty} T(k)C_{r+k-1}^{k}\theta^k(1-\theta)^r = \theta^s \quad \forall\theta \in (0, 1)$$

can be written as

$$\sum_{k=0}^{\infty} T(k)C_{r+k-1}^{k}\theta^k = \frac{\theta^s}{(1-\theta)^r} = \sum_{j=0}^{\infty} C_{r+j-1}^{j}\theta^{s+j} \quad \forall\theta \in (0, 1).$$

Since the two power series are identically equal, their respective coefficients are equal, and

$$T(k) = C_{r+k-s-1}^{r-1}/C_{r+k-1}^{r-1}, \quad k = 0, 1, \ldots,$$

is the only function which meets this identity. We conclude that the only unbiased estimator in this problem is the statistic

$$T(X) = (X)_s/(X + r - 1)_s.$$

If $r = 1$, this statistic only assumes the values 0 and 1 which do not belong to the parameter set $\Theta = (0, 1)$ of the model, and therefore it is practically useless.

2.13. Since $\mathscr{L}_\theta(\overline{X}) = \mathscr{N}(\theta, \sigma^2/n)$, we have $\mathbf{E}_\theta(\overline{X}^2) = \mathbf{D}_\theta\overline{X} + (\mathbf{E}_\theta\overline{X})^2 = (\sigma^2/n) + \theta^2$, whence follows the required assertion.

2.14. The second and fourth central moments of the distribution $\mathscr{N}(\mu, \theta^2)$ are $\mu_2 = \theta^2$ and $\mu_4 = \dfrac{4!}{2!2^2}\theta^4 = 3\theta^4$, respectively. Therefore (see the solution to Problem 2.1), we have

$$\mathbf{E}_\theta S^2 = \frac{n-1}{n}\theta^2, \quad \mathbf{D}_\theta S^2 = \frac{2(n-1)}{n^2}\theta^4.$$

Then

$$\mathbf{E}_\theta(S^2 - \theta^2)^2 = \mathbf{E}_\theta\left(\left(S^2 - \frac{n-1}{n}\theta^2\right) - \frac{\theta^2}{n}\right)^2$$

$$= \mathbf{D}_\theta(S^2) + \frac{\theta^4}{n} = \frac{2n-1}{n^2}\theta^4,$$

$$\mathbf{E}_\theta(\tau^* - \theta^2)^2 = \mathbf{D}_\theta\tau^* = \frac{1}{n}\mathbf{D}_\theta(X_1 - \mu)^2 = \frac{\mu_4 - \mu_2^2}{n} = \frac{2}{n}\theta^4,$$

$$\mathbf{D}_\theta(S'^2) = \left(\frac{n}{n-1}\right)^2\mathbf{D}_\theta(S^2) = \frac{2}{n-1}\theta^4.$$

Thus,

$$\mathbf{E}_\theta(S^2 - \theta^2)^2 < \mathbf{D}_\theta\tau^* < \mathbf{D}_\theta(S'^2),$$

i.e., by the minimal mean-square-error test the statistic S^2 is a more exact estimator for the theoretical variance θ^2 compared to the statistic τ^*, but in the class of unbiased estimators τ^* is more exact than S'^2.

2.15. Since $\mathscr{L}_\theta\left(\dfrac{X_i - \mu}{\theta}\right) = \mathscr{N}(0, 1)$, we have

$$\mathbf{E}_\theta|X_i - \mu| = \theta\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}|x|e^{-x^2/2}dx = \theta\sqrt{\frac{2}{\pi}}\int_0^{\infty}xe^{-x^2/2}dx = \theta\sqrt{\frac{2}{\pi}},$$

$$\mathbf{D}_\theta|X_i - \mu| = \mathbf{E}_\theta(X_i - \mu)^2 - (\mathbf{E}_\theta|X_i - \mu|)^2 = \frac{\pi - 2}{\pi}\theta^2.$$

Then $\mathbf{E}_\theta T_n(\mathbf{X}) = \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}_\theta|X_i - \mu| = \theta$ proves the unbiasedness and $\mathbf{D}_\theta T_n(\mathbf{X}) = \frac{\pi}{2}\frac{1}{n}\mathbf{D}_\theta|X_1 - \mu| = \frac{\pi - 2}{2n}\theta^2 = O\left(\frac{1}{n}\right)$ proves the consistency.

2.16. The formula $\mathscr{L}_\theta(T^2/\theta^2) = \chi^2(n) = \Gamma(2, n/2)$ and the formula for the moments of the gamma distribution imply that

$$\mathbf{E}_\theta T^k = \theta^k 2^{k/2}\Gamma\left(\frac{n+k}{2}\right)\Big/\Gamma\left(\frac{n}{2}\right),$$

i.e., the indicated estimator is unbiased.

In order to compare the estimator τ_1^* for θ with the estimator T_n from the previous problem, we must calculate $\mathbf{D}_\theta\tau_1^* = \mathbf{E}_\theta(\tau_1^*)^2 - \theta^2$. We have

$$\mathbf{E}_\theta(\tau_1^*)^2 = \frac{\Gamma^2\left(\frac{n}{2}\right)}{2\Gamma^2\left(\frac{n+1}{2}\right)}\mathbf{E}_\theta T^2 = \frac{n\Gamma^2\left(\frac{n}{2}\right)}{2\Gamma^2\left(\frac{n+1}{2}\right)}\theta^2$$

taking into account that $\Gamma(x+1) = x\Gamma(x)$. Consequently, $\mathbf{D}_\theta\tau_1^* = \left(\frac{n}{2}\frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n+1}{2}\right)} - 1\right)\theta^2$. This expression must be compared with $\mathbf{D}_\theta T_n = \frac{\pi - 2}{2n}\theta^2$ (see the solution to Problem 2.15). For $n = 1$ both statistics (and their variances) coincide. For $n = 2$ we have $\mathbf{D}_\theta\tau_1^* = \left(\frac{4}{\pi} - 1\right)\theta^2 \simeq 0.273\theta^2$ and $\mathbf{D}_\theta T_2 = \left(\frac{\pi}{4} - \frac{1}{2}\right)\theta^2 \simeq 0.285\theta^2$, i.e., the estimator τ_1^* is more exact than T_2 (here we have taken into account that $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$). Similarly, for $n = 3$ we have $\mathbf{D}_\theta\tau_1^* = \left(\frac{3\pi}{8} - 1\right)\theta^2 \simeq 0.178\theta^2 < \mathbf{D}_\theta T_3 \simeq 0.190\theta^2$. The result obtained in Problem 2.64 shows that the estimator τ_1^* is more exact than T_n for any n.

2.18. The mean square error of an arbitrary estimator T_λ is

$$\begin{aligned}\mathbf{E}_\theta(T_\lambda - \theta_2^2)^2 &= \mathbf{E}_\theta(\lambda(S'^2 - \theta_2^2) + (\lambda - 1)\theta_2^2)^2 \\ &= \lambda^2\mathbf{D}_\theta(S'^2) + (\lambda - 1)^2\theta_2^4 = \varphi(\lambda)\theta_2^4\end{aligned}$$

(see the solution to Problem 2.14, where $\mathbf{D}_\theta(S'^2)$ is calculated), and we have

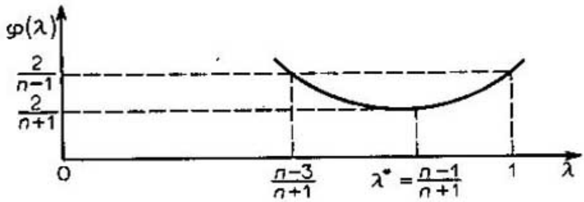


Fig. 7

$\varphi(\lambda) = \dfrac{2\lambda^2}{n-1} + (\lambda - 1)^2$. The function $\varphi(\lambda)$ is plotted in Fig. 7. Since for $\dfrac{n-3}{n+1} < \lambda < 1$ we have $\varphi(\lambda) < \varphi(1)$, the inequality

$$\mathbf{E}_\theta(T_\lambda - \theta_2^2)^2 < \mathbf{E}_\theta(S'^2 - \theta_2^2)^2$$

holds for these λ. We define k from the condition $\dfrac{n-3}{n+1} < \dfrac{n-1}{n+k} < 1$, which only holds for $k = 0, 1, 2, 3$. Since $\min\limits_\lambda \mathbf{E}_\theta(T_\lambda - \theta_2^2)^2 = \varphi(\lambda^*)\theta_2^4 = \dfrac{2}{n+1}\theta_2^4$, we conclude that the estimator $T_{\lambda^*} = \dfrac{1}{n+1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$ is optimal with respect to the minimal mean-square-error test.

2.19. From the solution to Problem 1.45 we have

$$\mathbf{E}_\theta T_\lambda = \lambda\theta_2^2, \quad \mathbf{E}_\theta T_\lambda^2 = \lambda^2 \frac{n+1}{n-1}\theta_2^4,$$

$$\mathbf{E}_\theta T_\lambda^3 = \lambda^3 \frac{(n+1)(n+3)}{(n-1)^2}\theta_2^6, \quad \mathbf{E}_\theta T_\lambda^4 = \lambda^4 \frac{(n+1)(n+3)(n+5)}{(n-1)^3}\theta_2^8.$$

We find the measure

$$\delta_1(\theta) = \mathbf{E}_\theta(T_\lambda - \theta_2^2)^4 = \mathbf{E}_\theta(T_\lambda^4 - 4T_\lambda^3\theta_2^2 + 6T_\lambda^2\theta_2^4 - 4T_\lambda\theta_2^6 + \theta_2^8) = \psi(\lambda)\theta_2^8,$$

where

$$\psi(\lambda) = \lambda^4 \frac{(n+1)(n+3)(n+5)}{(n-1)^3} - 4\lambda^3 \frac{(n+1)(n+3)}{(n-1)^2} + 6\lambda^2 \frac{n+1}{n-1} - 4\lambda + 1.$$

Using the formula $\lambda = \frac{n-1}{n+5}\left(1 + \frac{x}{n}\right)$, we reduce the equation $\psi'(\lambda) = 0$ to the form

$$x^3 + 3p_n x + 2q_n = 0, \quad p_n = \frac{2n^2}{n+3}, \quad q_n = -\frac{8n^3}{(n+1)(n+3)}.$$

The discriminant $D_n = p_n^3 + q_n^2$ of this equation is positive and has the only real root

$$x_n = \sqrt[3]{-q_n + \sqrt{D_n}} + \sqrt[3]{-q_n - \sqrt{D_n}}$$

(*Cardan's formula*), which has an asymptotic representation $x_n = \frac{8}{3} + O\left(\frac{1}{\sqrt{n}}\right)$ for large n.

We thus find the estimator $T^* = \frac{n-1}{n+5}\left(1 + \frac{x_n}{n}\right)S'^2$ which minimizes the measure $\delta_1(\theta)$ in the class of statistics $\{T_\lambda = \lambda S'^2\}$.

We now turn to the second measure

$$\delta_2(\theta) \equiv \mathbf{E}_\theta|T_\lambda - \theta_2^2| = \chi(\lambda)\theta_2^2,$$

where

$$\chi(\lambda) = \mathbf{E}_\theta\left|\frac{\lambda}{n-1}\chi_{n-1}^2 - 1\right|, \quad \mathscr{L}(\chi_{n-1}^2) = \chi^2(n-1).$$

If $k_{n-1}(t)$ is the density of the distribution $\chi^2(n-1)$, then

$$\chi(\lambda) = \int_0^\infty \left|\frac{\lambda}{n-1}t - 1\right| k_{n-1}(t)\,dt = \int_0^{\frac{n-1}{\lambda}} \left(1 - \frac{\lambda}{n-1}t\right) k_{n-1}(t)\,dt$$
$$+ \int_{\frac{n-1}{\lambda}}^\infty \left(\frac{\lambda}{n-1}t - 1\right) k_{n-1}(t)\,dt,$$

whence follows that the equation $\chi'(\lambda) = 0$ is equivalent to

$$\int_0^{\frac{n-1}{\lambda}} t k_{n-1}(t)\,dt = \int_{\frac{n-1}{\lambda}}^\infty t k_{n-1}(t)\,dt,$$

which defines the unique value $\lambda^* = \lambda_n^*$ and the optimum estimator T_{λ^*}.

2.20. Since $\mathscr{L}_\theta(nS^2/\theta_2^2) = \chi^2(n-1) = \Gamma\left(2, \frac{n-1}{2}\right)$, from the formula for the moments of the gamma distribution we have

$$\mathbf{E}_\theta S^k = \frac{\theta_2^k}{n^{k/2}}\mathbf{E}_\theta\left(\frac{n}{\theta_2^2}S^2\right)^{k/2} = \frac{\theta_2^k}{n^{k/2}}2^{k/2}\Gamma\left(\frac{n+k-1}{2}\right)\Big/\Gamma\left(\frac{n-1}{2}\right),$$

which means that the estimator is unbiased. For $n = 2$ we have $S = (1/2)|X_1 - X_2|$ and, therefore,

$$\tau_k^* = \frac{\sqrt{\pi}}{2^k\Gamma\left(\frac{k+1}{2}\right)}|X_1 - X_2|^k,$$

whence

$$\mathbf{E}_\theta|X_1 - X_2| - \theta_2 = \frac{2-\sqrt{\pi}}{\sqrt{\pi}}\theta_2.$$

2.21. The assertion immediately follows from the fact that $\mathscr{L}_\theta(T) = \Gamma(\theta, \lambda n)$ and from the formula for the moments of the gamma distribution.

2.22. If $\mathscr{L}_\theta(\xi) = \Gamma(\theta, 1)$, then $\mathscr{L}_\theta(\xi/\theta) = \Gamma(1, 1)$ and, according to Problem 1.34, the random variables $Y_r = \frac{n-r+1}{\theta}(X_{(r)} - X_{(r-1)})$, $r = 1, \ldots, n$, are independent, and we have $\mathscr{L}_\theta(Y_r) = \Gamma(1, 1)$ for any r. Then (see the solution to Problem 1.34) $X_{(k)} = \theta \sum_{j=1}^{k} Y_j/(n-j+1)$ and, therefore,

$$T = T(\mathbf{X}) = \sum_{k=1}^{r}\lambda_k X_{(k)} = \theta\sum_{i=1}^{r}\frac{\Lambda_i}{n-i+1}Y_i, \quad \Lambda_i = \sum_{k=i}^{r}\lambda_k,$$
$$i = 1, \ldots, r.$$

From this representation we immediately obtain

$$\mathbf{E}_\theta T = \theta\sum_{i=1}^{r}\Lambda_i/(n-i+1), \quad \mathbf{D}_\theta T = \theta^2\sum_{i=1}^{r}\Lambda_i^2/(n-i+1)^2.$$

The unbiasedness condition is equivalent to $\sum_{i=1}^{r}\Lambda_i/(n-i+1) = 1$, for which we have to minimize the expression $\sum_{i=1}^{r}\Lambda_i^2/(n-i+1)^2$. By applying the method of Lagrange multipliers, we find that the optimum choice of Λ_i will be $\Lambda_i^* = \frac{n-i+1}{r}$, $i = 1, \ldots, r$. We conclude that the optimum unbiased

estimator for θ has the form

$$T^* = \frac{\theta}{r}\sum_{i=1}^{r} Y_i = \frac{1}{r}\sum_{i=1}^{r}(n - i + 1)(X_{(i)} - X_{(i-1)}) = \frac{1}{r}\sum_{i=1}^{r} X_{(i)} + \frac{n-r}{r}X_{(r)},$$

and its variance is $\mathbf{D}_\theta T^* = \theta^2/r$.

2.23. It follows from the solution to Problem 1.36 that

$$\mathbf{E}_\theta T = \alpha\mathbf{E}_\theta X_{(n)} + \beta\mathbf{E}_\theta X_{(1)} = \left(\alpha\frac{2n+1}{n+1} + \beta\frac{n+2}{n+1}\right)\theta,$$

$$\mathbf{D}_\theta T = \alpha^2\mathbf{D}_\theta X_{(n)} + \beta^2\mathbf{D}_\theta X_{(1)} + 2\alpha\beta \operatorname{cov}(X_{(n)}, X_{(1)})$$

$$= \frac{n\theta^2}{(n+1)^2(n+2)}\left(\alpha^2 + \beta^2 + \frac{2\alpha\beta}{n}\right).$$

This implies that the optimum values of α and β must minimize the form $\alpha^2 + \beta^2 + 2\alpha\beta/n$ under the condition $\alpha(2n+1)/(n+1) + \beta(n+2)/(n+1) = 1$. The solution of this extremal problem (for example, using the Lagrange multipliers) has the form $\alpha^* = \dfrac{2(n+1)}{5n+4}$, $\beta^* = \dfrac{n+1}{5n+4}$.

Thus, $T^* = \dfrac{n+1}{5n+4}(X_{(1)} + 2X_{(n)})$ is an optimum unbiased estimator for θ in the class under consideration, and its variance is $\mathbf{D}_\theta T^* = \dfrac{\theta^2}{(n+2)(5n+4)}$.

2.24. The unbiasedness of these estimators directly follows from Problem 1.36. We have $\mathbf{D}_\theta T_1 = \dfrac{\theta^2}{n(n+2)} < \mathbf{D}_\theta T_2 = \dfrac{n}{n+2}\theta^2$, and therefore the estimator T_1 is more exact. Moreover, $\mathbf{D}_\theta T_1 \to 0$ as $n \to \infty$, i.e., T_1 is a consistent estimator, while T_2 does not have this property. Indeed, since

$$\mathbf{P}_\theta(X_{(1)} \leqslant t) = 1 - \left(1 - \frac{t}{\theta}\right)^n, \quad 0 \leqslant t \leqslant \theta,$$

we have

$$\mathbf{P}_\theta(|T_2 - \theta| \leqslant \varepsilon) = \mathbf{P}_\theta\left(\frac{\theta - \varepsilon}{n+1} \leqslant X_{(1)} \leqslant \frac{\theta + \varepsilon}{n+1}\right)$$

$$= \left(1 - \frac{\theta - \varepsilon}{\theta(n+1)}\right)^n - \left(1 - \frac{\theta + \varepsilon}{\theta(n+1)}\right)^n \to e^{-(\theta - \varepsilon)/\theta} - e^{-(\theta + \varepsilon)/\theta} < 1$$

as $n \to \infty$.

2.25. The formulas from Problem 1.36 directly show that the estimators are unbiased and give their variances, viz.,

$$\mathbf{D}_\theta T_1 = \frac{1}{4}(\mathbf{D}_\theta X_{(1)} + \mathbf{D}_\theta X_{(n)} + 2\operatorname{cov}(X_{(1)}, X_{(n)})) = \frac{(\theta_2-\theta_1)^2}{2(n+1)(n+2)},$$

$$\mathbf{D}_\theta T_2 = \left(\frac{n+1}{n-1}\right)^2(\mathbf{D}_\theta X_{(1)} + \mathbf{D}_\theta X_{(n)} - 2\operatorname{cov}(X_{(1)}, X_{(n)})) = \frac{2(\theta_2-\theta_1)^2}{(n-1)(n+2)}.$$

As $n \to \infty$ the variances tend to zero, which means that both estimators are consistent.

2.26. The results of Problem 1.37 directly show that the estimators are unbiased. Since $\mathbf{D}_\theta T \to 0$ as $n \to \infty$, they are also consistent.

2.27. In this case the theoretical mean coincides with θ, and the solution to Problem 2.1 (b) gives the required result.

2.28. By the property of the arithmetic mean of Cauchy's distribution we have $\mathscr{L}_\theta(\overline{X}) = C(\theta)$, i.e., the distribution of the statistic $\overline{X}$ is independent of n, and therefore the quantity $\mathbf{P}_\theta(|\overline{X}-\theta| \geqslant \varepsilon)$ is the same for all n.

2.29. It follows from Problem 1.52 that as $n \to \infty$

$$\mathbf{E}_\theta T_r = p_r,\ \mathbf{D}_\theta T_r = \frac{1}{n^2}[\mathbf{E}_\theta \nu_r(\nu_r - 1) + \mathbf{E}_\theta \nu_r - (\mathbf{E}_\theta \nu_r)^2] = \frac{p_r(1-p_r)}{n} \to 0$$

and this proves the assertion (a). We then have

$$\mathbf{E}_\theta H(T_1, \ldots, T_N) = \sum_{k_1+\ldots+k_N=n} H\left(\frac{k_1}{n}, \ldots, \frac{k_N}{n}\right)\frac{n!}{k_1!\ldots k_N!}p_1^{k_1}\ldots p_N^{k_N}.$$

For any function H the right-hand side of this is a polynomial in $p_1, \ldots, p_N$ of degree $\leqslant n$. Consequently, the unbiased estimators can only be constructed for the polynomials of degree $\leqslant n$ in the parameters $p_1, \ldots, p_N$.

Finally, if $H = \dfrac{1}{n}\sum\limits_{i=1}^{N} c_i\nu_i$, then

$$\mathbf{E}_\theta H = \sum_{i=1}^{N} c_i p_i = \tau(\theta),\quad \mathbf{D}_\theta H = \frac{1}{n}\left(\sum_{i=1}^{N} c_i^2 p_i - \tau^2(\theta)\right) \to 0$$

as $n \to \infty$, i.e., H is an unbiased and consistent estimator for $\tau(\theta)$.

2.30. We have $\alpha_1 = \alpha_1(\theta) = \theta_1\Gamma(\theta_2+1)/\Gamma(\theta_2) = \theta_1\theta_2$, $\alpha_2 = \alpha_2(\theta) = \theta_1^2\Gamma(\theta_2+2)/\Gamma(\theta_2) = \theta_1^2\theta_2(\theta_2+1)$, and $\theta_1 = (\alpha_2 - \alpha_1^2)/\alpha_1$, $\theta_2 = \alpha_1^2/(\alpha_2-\alpha_1^2)$. The sought-for estimates have the form

$$\tilde\theta_1 = (A_{n2} - A_{n1}^2)/A_{n1} = S^2/\overline{X},$$

$$\tilde\theta_2 = A_{n1}^2/(A_{n2} - A_{n1}^2) = \overline{X}^2/S^2.$$

These statistics are continuous functions of the sampling moments, and therefore they are consistent estimates for the respective parameters.

2.31. Here $\alpha_1(\theta) = \mathbf{E}_\theta\xi = \frac{1}{2}(\theta_1 + \theta_2)$, $\alpha_2(\theta) = \mathbf{E}_\theta\xi^2 = \frac{1}{2}(\theta_1^2 + \theta_2^2 + \theta_1 + \theta_2)$, and the estimates

$$\tilde{\theta}_{1,2} = A_{n1} \mp \sqrt{A_{n2} - A_{n1}^2 - A_{n1}} = \overline{X} \mp \sqrt{S^2 - \overline{X}}$$

are the solutions to the equations $\alpha_k(\theta) = A_{nk}$, $k = 1, 2$. For the indicated data we have $\tilde{\theta}_1 = 2.17\ldots$, $\tilde{\theta}_2 = 3.57\ldots$.

2.33. If $D \leqslant k_0$, then the unbiasedness conditions are equivalent to the system

$$\sum_{k=0}^{D} T(k)f(k;\, D,\, n) = \tau(D), \quad D = 0, 1, \ldots, k_0.$$

This is a triangular system with the diagonal coefficients $f(D;\, D,\, n) = C_{N-D}^{n-D}/C_N^n \neq 0$ and the values $T(0), T(1), \ldots, T(k_0)$ are uniquely defined from it. For the values $m > k_0$ we stipulate that

$$T(m) = \left[\tau(m) - \sum_{k=0}^{k_0} T(k)f(k;\, m,\, n)\right] \Big/ \left[1 - \sum_{k=0}^{k_0} f(k;\, m,\, n)\right],$$

which can be done because for $m > k_0$ we have $1 - \sum_{k=0}^{k_0} f(k;\, m,\, n) \neq 0$.

Since

$$\sum_{k=0}^{D} kf(k;\, D,\, n) = \frac{Dn}{N}$$

(see the formula for the mean of the hypergeometric distribution $H(D, N, n)$ in Chap. 1), for the case of $\tau(D) = D$ the function T has the form $T(k) = kN/n$ for $k = 0, 1, \ldots, k_0$, and for $m > k_0$ we have

$$T(m) = \left[m - \frac{N}{n}\sum_{k=0}^{k_0} kf(k;\, m,\, n)\right] \Big/ \left[1 - \sum_{k=0}^{k_0} f(k;\, m,\, n)\right].$$

Specifically, if we put $k_0 = n$ (the inspection of the entire batch is not carried out), then the statistic $T(\xi) = \frac{N}{n}\xi$ is an unbiased estimator for the number D of the defective items.

2.34. (a) Since $\mathbf{E}\gamma(u) = \mathbf{P}(u \in s) = \pi(u)$ (see the hint), the representation $\bar{e}(s, \mathbf{X}) = \sum_u \gamma(u)x(u)/\pi(u)$ implies that the Horvitz-Thompson estimator is unbiased. Since $\sum_{u \in s} a(u)x(u) = \sum_u \gamma(u)a(u)x(u)$, the unbiasedness condi-

tion for the estimator implies that

$$\sum_u \pi(u)a(u)x(u) = \sum_u x(u) \quad \forall \mathbf{x} \in R^N.$$

Specifically, if we take the coordinate vectors of the Euclidean space R^N as $\mathbf{x}$, we will find that $\pi(u_i)a(u_i) = 1$, $i = 1, \ldots, N$. Thus, the Horvitz-Thompson estimator is the only linear unbiased estimator for $T(\mathbf{x})$.

(b) Since

$$\bar{e}^2(s, \mathbf{x}) = \sum_{u \neq v} \gamma(u)\gamma(v)x(u)x(v)/\pi(u)\pi(v) + \sum_u \gamma(u)x^2(u)/\pi^2(u)$$

and

$$\mathbf{E}\gamma(u)\gamma(v) = \mathbf{P}(u \in s,\ v \in s) = \pi(u, v), \quad u \neq v,$$

we have

$$\mathbf{D}\bar{e}(s, \mathbf{x}) = \mathbf{E}\bar{e}^2(s, \mathbf{x}) - (\mathbf{E}\bar{e}(s, \mathbf{x}))^2 = \sum_{u \neq v} \frac{\pi(u, v)}{\pi(u)\pi(v)} x(u)x(v)$$

$$+ \sum_u x^2(u)/\pi(u) - \left(\sum_u x(u)\right)^2,$$

which is equivalent to the required formula.

(c) The unbiasedness follows from the representation

$$\Delta(s, \mathbf{x}) = \sum_u \gamma(u) \frac{x^2(u)}{\pi(u)} \left(\frac{1}{\pi(u)} - 1\right)$$

$$+ \sum_{u \neq v} \gamma(u)\gamma(v) \frac{x(u)x(v)}{\pi(u, v)} \left(\frac{\pi(u, v)}{\pi(u)\pi(v)} - 1\right).$$

(d) The formulas directly follow from

$$n(s) = \sum_u \gamma(u), \quad n^2(s) = \sum_{u \neq v} \gamma(u)\gamma(v) + \sum_u \gamma(u).$$

2.35. (a) For any fixed unit u there is $n(N - 1)_{n-1}$ different samples which contain this element. Therefore,

$$\pi(u) = \frac{n(N - 1)_{n-1}}{(N)_n} = \frac{n}{N}.$$

(b) The formula for $\mathbf{D}\bar{x}$ follows from the general formula for $\mathbf{D}\bar{e}(s, \mathbf{x})$ (see Problem 2.34 (b)) if we take into account that

$$\pi(u, v) = \frac{n(n-1)(N-2)_{n-2}}{(N)_n} = \frac{n(n-1)}{N(N-1)}, \quad u \neq v.$$

(c) The unbiasedness of $\hat{\sigma}^2(s, \mathbf{x})$ follows from

$$\hat{\sigma}^2(s, \mathbf{x}) = \frac{1}{n} \sum_u \gamma(u)(x(u) - \mu)^2$$

$$- \frac{1}{n(n-1)} \sum_{u \neq v} \gamma(u)\gamma(v)(x(u) - \mu)(x(v) - \mu).$$

2.36. We first calculate $\mathbf{E}\mu_r$. We use the hint to obtain $\mathbf{E}\mu_r = N\mathbf{E}\xi_1^{(r)} = N\mathbf{P}(\xi_1^{(r)} = 1)$, where, according to the classical definition of probability, we have

$$\mathbf{P}(\xi_1^{(r)} = 1) = C_n^r (C_{N-1}^{m-1})^r (C_{N-1}^m)^{n-r} / (C_N^m)^n = C_n^r \left(\frac{m}{N}\right)^r \left(1 - \frac{m}{N}\right)^{n-r}.$$

Finally,

$$\mathbf{E}\mu_r = NC_n^r \left(\frac{m}{N}\right)^r \left(1 - \frac{m}{N}\right)^{n-r}.$$

We find that the mean of any linear statistic has the form

$$\mathbf{E}l = N \sum_{r=1}^{n} l_r C_n^r \left(\frac{m}{N}\right)^r \left(1 - \frac{m}{N}\right)^{n-r},$$

i.e., it is a polynomial in $1/N$ of degree $\leqslant n - 1$. This means that in the class $\mathscr{L}$ the unbiased estimators can only exist for the parametric functions of the form $\tau(N) = \sum_{j=1}^{k} c_j/N^j$ for $k \leqslant n - 1$. Let $\tau(N)$ be an arbitrary function of this kind. Then the unbiasedness condition implies that

$$\sum_{r=1}^{n} l_r C_n^r \left(\frac{m}{N}\right)^r \left(1 - \frac{m}{N}\right)^{n-r} = \sum_{j=1}^{k} c_j/N^{j+1} \quad \forall N \geqslant m.$$

It follows that the coefficients l_r of the sought-for estimator are uniquely defined through the coefficients c_j. By taking into account that

$$\sum_{r=0}^{n} (r)_j C_n^r \left(\frac{m}{N}\right)^r \left(1 - \frac{m}{N}\right)^{n-r} = (n)_j \left(\frac{m}{N}\right)^j, \quad j = 1, 2, \ldots,$$

we can directly check that the coefficients l_r have the required form (see Problem 1.52 (b)).

2.37. We first find the distribution of the random variable η. We use A_i to denote the event consisting in that the ith element ($i = 1, \ldots, N$) has not been observed, and suppose that $\mu_0(n, m, N) = N - \eta$ is the total number of the elements which have not been observed. Then, by the formula for the sum of probabilities [2], we will have

$$\mathbf{P}(\mu_0(n, m, N) > 0) = \mathbf{P}\left(\bigcup_{i=1}^{N} A_i\right)$$

$$= \sum_{j=1}^{N} (-1)^{j+1} \sum_{1 \leqslant k_1 < \ldots < k_j \leqslant N} \mathbf{P}(A_{k_1} \ldots A_{k_j}).$$

In the internal sum all the terms are $(C_{N-j}^m)^n/(C_N^m)^n$, and their number is C_N^j. Therefore,

$$\mathbf{P}(\mu_0(n, m, N) > 0) = \sum_{j=1}^{N} (-1)^{j+1} C_N^j (C_{N-j}^m)^n/(C_N^m)^n.$$

Then

$$\mathbf{P}(\eta = k) = \mathbf{P}(\mu_0(n, m, N) = N - k)$$

$$= \sum_{1 \leqslant i_1 < \ldots < i_{N-k} \leqslant N} \mathbf{P}(A_{i_1} \ldots A_{i_{N-k}} \overline{A}_{j_1} \ldots \overline{A}_{j_k}),$$

where $\{j_1, \ldots, j_k\} = \{1, \ldots, N\} \setminus \{i_1, \ldots, i_{N-k}\}$. In the latter sum all the terms are equal, and their number is C_N^k. By the theorem on the product of probabilities, we will have

$$\mathbf{P}(A_{i_1} \ldots A_{i_{N-k}} \overline{A}_{j_1} \ldots \overline{A}_{j_k}) = \mathbf{P}(A_{i_1} \ldots A_{i_{N-k}}) \mathbf{P}(\overline{A}_{j_1} \ldots \overline{A}_{j_k} | A_{i_1} \ldots A_{i_{N-k}}).$$

Here

$$\mathbf{P}(A_{i_1} \ldots A_{i_{N-k}}) = (C_k^m)^n/(C_N^m)^n$$

and

$$\mathbf{P}(\overline{A}_{j_1} \ldots \overline{A}_{j_k} | A_{i_1} \ldots A_{i_{N-k}}) = \mathbf{P}(\mu_0(n, m, k) = 0)$$

$$= \sum_{j=0}^{k} (-1)^j C_k^j (C_{k-j}^m)^n/(C_k^m)^n.$$

These relations give

$$\mathbf{P}(\eta = k) = C_N^k \sum_{j=0}^{k} (-1)^{k-j} C_k^j (C_j^m)^n/(C_N^m)^n,$$

$$k = m, m + 1, \ldots, \min(m, n, N).$$

Now let $N \leqslant mn$. Then

$$\mathbf{E}\tau^* = \sum_{k=m}^{N} \tau^*(k)\mathbf{P}(\eta = k) = (C_N^m)^{-n} \sum_{k=m}^{N} C_N^k \sum_{j=m}^{k} (-1)^{k-j} C_k^j (C_j^m)^n \tau(j)$$

$$= (C_N^m)^{-n} \sum_{j=m}^{N} \tau(j)(C_j^m)^n C_N^j \sum_{r=0}^{N-j} (-1)^r C_{N-j}^r = \tau(N)$$

because

$$\sum_{r=0}^{N-j} (-1)^r C_{N-j}^r = \begin{cases} 0 & \text{for } j < N, \\ 1 & \text{for } j = N. \end{cases}$$

If $N > mn$, then by taking into account the properties of the function $f(N)$, we may write

$$\mathbf{E}\tau^* = (C_N^m)^{-n} \sum_{k=0}^{N} C_N^k \sum_{j=0}^{k} (-1)^{k-j} C_k^j f(j).$$

It is sufficient to show that $\sum_{j=0}^{k} (-1)^{k-j} C_k^j f(j) = 0$ for $k > mn$, which follows from the chain of equations

$$\sum_{j=0}^{k} (-1)^{k-j} C_k^j (j)_r = (k)_r \sum_{s=0}^{k-r} (-1)^{k-r-s} C_{k-r}^s = 0, \quad r < k.$$

We now have

$$\mathbf{E}\tau^* = (C_N^m)^{-n} \sum_{j=0}^{N} f(j) C_N^j \sum_{r=0}^{N-j} (-1)^r C_{N-j}^r = (C_N^m)^{-n} f(N) = \tau(N).$$

2.41. Since in repeated independent trials the vectors $\mathbf{X}$ and $\pi\mathbf{X}$ are distributed in the same way, we have $\mathbf{E}_\theta T^* = \mathbf{E}_\theta T = \tau$, i.e., T^* is an unbiased estimator for $\tau(\theta)$. Now let $\mathbf{D}_\theta T(\mathbf{X}) = \delta^2$, then $\mathbf{D}_\theta T(\pi\mathbf{X}) = \delta^2$ and, according to the Cauchy-Schwarz inequality, we have

$$\operatorname{cov}_\theta (T(\pi_1\mathbf{X}),\ T(\pi_2\mathbf{X})) \leqslant \sqrt{\mathbf{D}_\theta T(\pi_1\mathbf{X})\mathbf{D}_\theta T(\pi_2\mathbf{X})} = \delta^2.$$

We then find

$$\mathbf{D}_\theta T^* = \frac{1}{(n!)^2}\left[\sum_{\pi} \mathbf{D}_\theta T(\pi\mathbf{X}) + \sum_{\pi_1 \neq \pi_2} \operatorname{cov}_\theta (T(\pi_1\mathbf{X}),\ T(\pi_2\mathbf{X}))\right]$$

$$\leqslant \delta^2 \frac{n! + n!(n! - 1)}{(n!)^2} = \delta^2.$$

We see that for any unbiased estimator we can find a symmetric one whose variance is not greater than the variance of the original estimate. Consequently, the optimum estimator (if it exists) should be sought among the symmetric functions of observations.

2.42. (1) Consider the statistic $T_\lambda = T^* + \lambda\psi$, which for any λ is an unbiased estimator for τ. Then, since T^* is optimal, we have

$$\mathbf{D}_\theta T_\lambda = \mathbf{D}_\theta T^* + \lambda^2 \mathbf{D}_\theta \psi + 2\lambda \operatorname{cov}_\theta (T^*, \psi) \geqslant \mathbf{D}_\theta T^*.$$

But this is possible for all λ if only $\operatorname{cov}_\theta (T^*, \psi) = 0 \quad \forall\theta$.

(2) Let T be an arbitrary unbiased estimator for τ. Then the statistic $\psi = T^* - T$ has a zero mean, and we find from the above that

$$0 = \operatorname{cov}_\theta (T^*, T^* - T) = \mathbf{D}_\theta T^* - \operatorname{cov}_\theta (T^*, T).$$

2.43. In the model $\mathcal{N}(\theta, \sigma^2)$ we have $f(x; \theta) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$. Then $-\dfrac{\partial^2}{\partial\theta^2} \ln f(x; \theta) = \dfrac{1}{\sigma^2}$. Therefore,

$$i(\theta) = \frac{1}{\sigma^2}.$$

In the model $\mathcal{N}(\mu, \theta^2)$ we have $-\dfrac{\partial^2}{\partial\theta^2} \ln f(x; \theta) = \dfrac{3(x-\mu)^2}{\theta^4} - \dfrac{1}{\theta^2}$, whence

$$i(\theta) = \frac{3}{\theta^4} \mathbf{E}_\theta (X_1 - \mu)^2 - \frac{1}{\theta^2} = \frac{2}{\theta^2}.$$

In the model $\Gamma(\theta, \lambda)$ we have $f(x; \theta) = \dfrac{x^{\lambda-1} e^{-x/\theta}}{\Gamma(\lambda)\theta^\lambda}$ and $-\dfrac{\partial^2}{\partial\theta^2} \ln f(x; \theta) = \dfrac{2x}{\theta^3} - \dfrac{\lambda}{\theta^2}$, whence

$$i(\theta) = \frac{2}{\theta^3} \mathbf{E}_\theta X_1 - \frac{\lambda}{\theta^2} = \frac{\lambda}{\theta^2}.$$

In Cauchy's model $\dfrac{\partial}{\partial\theta} \ln f(x; \theta) = \dfrac{2(x-\theta)}{1 + (x-\theta)^2}$. Therefore,

$$i(\theta) = \frac{4}{\pi} \int_{-\infty}^{\infty} \frac{(x-\theta)^2}{[1 + (x-\theta)^2]^3} dx = \frac{1}{2\pi} \arctan x \Big|_{-\infty}^{\infty} = \frac{1}{2}.$$

In the binomial model we have $f(x; \theta) = C_k^x \theta^x (1-\theta)^{k-x}$ and $-\dfrac{\partial^2}{\partial\theta^2} \ln f(x; \theta) = \dfrac{x}{\theta^2} + \dfrac{k-x}{(1-\theta)^2}$. Consequently,

$$i(\theta) = \frac{1}{\theta^2}\mathbf{E}_\theta X_1 + \frac{1}{(1-\theta)^2}(k - \mathbf{E}_\theta X_1) = \frac{k}{\theta} + \frac{k}{1-\theta} = \frac{k}{\theta(1-\theta)}.$$

In Poisson's model $f(x;\theta) = e^{-\theta}\dfrac{\theta^x}{x!}$ and $-\dfrac{\partial^2}{\partial\theta^2}\ln f(x;\theta) = \dfrac{x}{\theta^2}$. Consequently,

$$i(\theta) = \frac{1}{\theta^2}\mathbf{E}_\theta X_1 = \frac{1}{\theta}.$$

In the model $\overline{Bi}(r, \theta)$ we have $f(x;\theta) = C^x_{r+x-1}\theta^x(1-\theta)^r$ and $-\dfrac{\partial^2}{\partial\theta^2}\ln f(x;\theta) = \dfrac{x}{\theta^2} + \dfrac{r}{(1-\theta)^2}$. Consequently,

$$i(\theta) = \frac{r}{\theta(1-\theta)} + \frac{r}{(1-\theta)^2} = \frac{r}{\theta(1-\theta)^2}.$$

2.44. In this case we have

$$\ln f(x;\theta) = \frac{-(x-\theta_1)^2}{2\theta_2^2} - \ln(\theta_2\sqrt{2\pi})$$

and

$$-\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta_1\partial\theta_2}\ln f(X_1;\theta)\right) = \frac{2}{\theta_2^3}\mathbf{E}_\theta(X_1 - \theta_1) = 0.$$

This fact and the previous results give us the required formula.

2.45. We find the form of the matrix $\mathbf{I}(\theta)$ from the formulas

$$\mathbf{E}_\theta\delta(\xi, a_j) = \mathbf{P}_\theta(\xi = a_j) = p_j, \quad -\frac{\partial^2 \ln f(\xi, a_i)}{\partial p_i^2} = \frac{\delta(\xi, a_i)}{p_i^2} + \frac{\delta(\xi, a_N)}{p_N^2},$$

$$-\frac{\partial^2 \ln f(\xi, a_i)}{\partial p_i \partial p_j} = \frac{\delta(\xi, a_N)}{p_N^2}, \quad i \neq j.$$

We can directly check that $\mathbf{I}^{-1}(\theta) = \Sigma_{N-1}$, where the matrix Σ_{N-1} is defined as in Problem 1.53.

2.46. Suppose that $\mathcal{F}$ is an exponential model. Then for $\theta = (\theta_1, \ldots, \theta_r)$ we have

$$U_j(\mathbf{X};\theta) = \frac{\partial}{\partial\theta_j}\ln L(\mathbf{X};\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta_j}\ln f(X_i;\theta)$$

$$= \frac{\partial A(\theta)}{\partial\theta_j}T(\mathbf{X}) + n\frac{\partial C(\theta)}{\partial\theta_j},$$

where $T(\mathbf{X}) = \sum_{i=1}^{n} B(X_i)$. We take $a_j(\theta) = \left(n \frac{\partial A(\theta)}{\partial \theta_j}\right)^{-1}$, $j = 1, \ldots, r$, and write

$$\sum_{j=1}^{r} a_j(\theta) U_j(\mathbf{X};\ \theta) = \frac{r}{n} T(\mathbf{X}) + \sum_{j=1}^{r} \frac{\partial C(\theta)}{\partial \theta_j} \Bigg/ \frac{\partial A(\theta)}{\partial \theta_j} = \tau^* - \tau(\theta).$$

Conversely, if we have a representation $\mathbf{a}'(\theta)\mathbf{U}(\mathbf{X};\ \theta) = T_n(\mathbf{X}) - \tau(\theta)$ for some $\mathbf{a}(\theta) = (a_1(\theta), \ldots, a_r(\theta))$, $T_n(\mathbf{X})$, and $\tau(\theta)$, then we have a special case of

$$\sum_{j=1}^{r} a_j(\theta) \frac{\partial \ln f(x;\ \theta)}{\partial \theta_j} = T_1(x) - \tau(\theta).$$

This means that the function $f(\mathbf{x};\ \theta)$ has the required form.

In order to obtain the variance $\mathbf{D}_\theta \tau^*$, we recall that

$$\begin{aligned} \frac{\partial \tau(\theta)}{\partial \theta_j} &= \int \tau^*(\mathbf{x}) \frac{\partial \ln L(\mathbf{x};\ \theta)}{\partial \theta_j} L(\mathbf{x};\ \theta)\, d\mathbf{x} \\ &= \mathbf{E}_\theta(\tau^*(\mathbf{X}) U_j(\mathbf{X};\ \theta)) = \operatorname{cov}_\theta(\tau^*(\mathbf{X}),\ U_j(\mathbf{X};\ \theta)) \end{aligned}$$

because $\mathbf{E}_\theta U_j(\mathbf{X};\ \theta) = 0\ \forall \theta$. Therefore,

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^{r} \frac{\partial \tau(\theta)}{\partial \theta_j} \Bigg/ \frac{\partial A(\theta)}{\partial \theta_j} &= \operatorname{cov}_\theta(\tau^*,\ \mathbf{a}'(\theta)\mathbf{U}(\mathbf{X};\ \theta)) \\ &= \operatorname{cov}_\theta(\tau^*,\ \tau^* - \tau(\theta)) = \mathbf{D}_\theta \tau^*. \end{aligned}$$

2.47. Since the variance of the efficient estimator coincides with the Cramér-Rao bound, we find from the previous problem that $\frac{\tau'(\theta)}{A'(\theta)} = \frac{[\tau'(\theta)]^2}{i(\theta)}$, which gives the required expression for $i(\theta)$. We also have

$$i(\theta) = -\mathbf{E}_\theta \frac{\partial^2 \ln f(\xi;\ \theta)}{\partial \theta^2} = -A''(\theta)\mathbf{E}_\theta B(\xi) - C''(\theta) = \frac{C'(\theta)}{A'(\theta)} A''(\theta) - C''(\theta),$$

i.e., $\mathbf{E}_\theta B(\xi) = -C'(\theta)/A'(\theta)$.

2.48. We directly apply the results of Problem 2.46. For example, in the model $\overline{Bi}(r,\ \theta)$ we have

$$f(x;\ \theta) = \exp\{x \ln \theta + r \ln(1 - \theta) + \ln C^x_{r+x-1}\},$$

i.e., $A(\theta) = \ln \theta$, $B(x) = x$, $C(\theta) = r \ln(1 - \theta)$. Consequently,

$$\tau(\theta) = -C'(\theta)/A'(\theta) = r\theta/(1 - \theta),\ \tau^* = \overline{X},\ \mathbf{D}_\theta \tau^* = \frac{\tau'(\theta)}{nA'(\theta)} = \frac{r\theta}{n(1 - \theta)^2}.$$

2.49. We have $-\dfrac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} = 2f(x;\theta)$, whence

$$i(\theta) = 2\int_{-\infty}^{\infty} f^2(x;\theta)\,dx = 2\int_0^{\infty}\frac{t\,dt}{(1+t)^4} = \frac{1}{3}.$$

Then

$$\mathbf{D}_\theta\xi = \mathbf{E}_\theta(\xi-\theta)^2 = \int_{-\infty}^{\infty}(x-\theta)^2 f(x;\theta)\,dx = \int_{-\infty}^{\infty} x^2 e^{-x}(1+e^{-x})^{-2}\,dx = \frac{\pi^2}{3}.$$

From this we find

$$\mathbf{D}_\theta\overline{X} = \frac{\pi^2}{3n} > \frac{1}{ni(\theta)} = \frac{3}{n}.$$

2.50. The assertion follows from

$$\frac{1}{L}\left(\frac{2\theta\sigma^2}{n}\frac{\partial L}{\partial\theta} + \frac{\sigma^4}{n^2}\frac{\partial^2 L}{\partial\theta^2}\right) = \bar{x}^2 - \frac{\sigma^2}{n} - \theta^2$$

and the Bhattacharyya test.

2.51. It follows from Problem 2.21 that $\mathbf{E}_\theta T^* = \theta^2$, $\mathbf{D}_\theta T^* = \dfrac{2(2\lambda n + 3)}{\lambda n(\lambda n + 1)}\theta^4$. The lower Cramér-Rao bound for the function $\tau(\theta) = \theta^2$ is (see Problem 2.43) $\dfrac{4}{\lambda n}\theta^4$, which is smaller than $\mathbf{D}_\theta T^*$. Therefore, T^* is not an efficient estimator for $\tau(\theta)$. The optimality of T^* follows from

$$\frac{1}{L}\left(\frac{2\theta^3}{\lambda n}\frac{\partial L}{\partial\theta} + \frac{\theta^4}{\lambda n(\lambda n+1)}\frac{\partial^2 L}{\partial\theta^2}\right) = \frac{n}{\lambda(\lambda n+1)}\bar{x}^2 - \theta^2$$

and the Bhattacharyya test.

2.52. The optimality of the estimators follows from

$$\frac{\theta_2^2}{n}\frac{\partial \ln L}{\partial\theta_1} = \bar{x} - \theta_1,$$

$$\frac{1}{L}\left(\frac{\theta_2^3}{n-1}\frac{\partial L}{\partial\theta_2} - \frac{\theta_2^4}{n(n-1)}\frac{\partial^2 L}{\partial\theta_1^2}\right) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \theta_2^2$$

and the Bhattacharyya test. The information matrix for the model $\mathcal{N}(\theta_1, \theta_2^2)$ was found in Problem 2.44, whence we find the Cramér-Rao bound θ_2^2/n for

the function $\tau_1(\theta) = \theta_1$, which coincides with $\mathbf{D}_\theta \overline{X}$, i.e., $\overline{X}$ is an efficient estimator for $\tau_1(\theta)$. For the function $\tau_2(\theta) = \theta_2^2$ the bound is $2\theta_2^4/n$, which is smaller than $\mathbf{D}_\theta(S'^2) = 2\theta_2^4/(n-1)$ (see the solution to Problem 2.14), i.e., S'^2 is not an efficient estimator for $\tau_2(\theta)$.

2.53. It is clear from the previous problem that the sample mean of the common sample is the best estimator for the mean, i.e., the statistic $\overline{X} = \frac{1}{n}(n_1\overline{X_1} + n_2\overline{X_2})$, where $n = n_1 + n_2$. We also have

$$\mathbf{D}_\theta \overline{X} = \frac{\theta_2^2}{n} < \min(\mathbf{D}_\theta \overline{X_1}, \mathbf{D}_\theta \overline{X_2}) = \theta_2^2/\max(n_1, n_2).$$

Similarly, the statistic

$$S'^2 = \frac{n}{n-1}(A_{n2} - \overline{X}^2)$$

is the best estimator for the variance and takes into account all the information. But $nA_{n2} = n_1 A_{n_1 2}^{(1)} + n_2 A_{n_2 2}^{(2)}$, where $A_{n_i 2}^{(i)}$ is the second-order sampling moment of the ith sample, $i = 1, 2$. From the formula

$$S_i'^2 = \frac{n_i}{n_i - 1}(A_{n_i 2}^{(i)} - \overline{X}_i^2)$$

we have $n_i A_{n_i 2}^{(i)} = (n_i - 1)S_i'^2 + n_i\overline{X_i^2}$. We finally find that

$$S'^2 = \frac{1}{n-1}((n_1 - 1)S_1'^2 + (n_2 - 1)S_2'^2 + n_1\overline{X}_1^2 + n_2\overline{X}_2^2 - n\overline{X}^2)$$

and (see Problem 2.14)

$$\mathbf{D}_\theta(S'^2) = \frac{2}{n-1}\theta_2^4 < \min(\mathbf{D}_\theta(S_1'^2), \mathbf{D}_\theta(S_2'^2)) = \frac{2\theta_2^4}{\max(n_1, n_2) - 1}.$$

2.54. (1) Suppose that λ is known. We are dealing with the exponential model with (see Problem 2.46)

$$B(x) = x,\ A(\theta) = -\frac{\lambda}{2\theta^2},\ C(\theta) = \frac{\lambda}{\theta},$$

where $\theta = \mu$. The efficient estimator here is $\tau^* = \overline{X}$, and the respective parametric function is $\tau(\theta) = \mu$. We also have $\mathbf{D}\tau^* = \frac{1}{n}\mathbf{D}X_1 = \frac{1}{nA'(\theta)} = \frac{\mu^3}{n\lambda}$. If λ is unknown, we calculate

$$\frac{\partial \ln L}{\partial \mu} = \frac{\lambda n}{\mu^3}(\overline{X} - \mu)$$

and apply the Bhattacharyya test to obtain the required result.

(2) We are dealing here with the exponential model

$$B(x) = \frac{x}{\mu^2} + \frac{1}{x},\ A(\theta) = -\frac{\theta}{2},\ C(\theta) = \frac{\theta}{\mu} + \frac{1}{2}\ln\theta,$$

where $\theta = \lambda$ and, according to the solution of Problem 2.46, the statistic $\tau^* = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i}{\mu^2} + \frac{1}{X_i}\right)$ is an efficient estimator for $\tau(\theta) = \frac{1}{\lambda} + \frac{2}{\mu}$. Consequently, the sought-for estimator has the form $\tau^* - 2/\mu$.

2.55. Using the Cramér-Rao inequality for scalar estimators for all $\mathbf{c} \in R^m$, we have

$$\mathbf{D}_\theta(\mathbf{c}'\mathbf{T}) \geqslant \mathbf{b}'(\theta)\mathbf{I}_n^{-1}(\theta)\mathbf{b}(\theta),$$

where

$$\mathbf{b}(\theta) = \left(\mathbf{c}'\frac{\partial\tau(\theta)}{\partial\theta_1}, \ldots, \mathbf{c}'\frac{\partial\tau(\theta)}{\partial\theta_r}\right)$$
$$= \left(\sum_{i=1}^{m} c_i \frac{\partial\tau_i(\theta)}{\partial\theta_1}, \ldots, \sum_{i=1}^{m} c_i \frac{\partial\tau_i(\theta)}{\partial\theta_r}\right) = \mathbf{B}(\theta)\mathbf{c}.$$

Thus,

$$\mathbf{c}'\mathbf{D}_\theta(\mathbf{T})\mathbf{c} \geqslant \mathbf{c}'\mathbf{B}'(\theta)\mathbf{I}_n^{-1}(\theta)\mathbf{B}(\theta)\mathbf{c}$$

or

$$\mathbf{c}'[\mathbf{D}_\theta(\mathbf{T}) - \mathbf{B}'(\theta)\mathbf{I}_n^{-1}(\theta)\mathbf{B}(\theta)]\mathbf{c} \geqslant 0.$$

This means that the matrix $\mathbf{D}_\theta(\mathbf{T}) - \mathbf{B}'(\theta)\mathbf{I}_n^{-1}(\theta)\mathbf{B}(\theta)$ is non-negative definite.

2.56. If an efficient estimator exists for $\tau(\theta)$, then the model is exponential (see Problem 2.46). Therefore,

$$L(\mathbf{x};\,\theta) = \exp\left\{A(\theta)T(\mathbf{x}) + nC(\theta) + \sum_{i=1}^{n} D(x_i)\right\},\ T(\mathbf{x}) = \sum_{i=1}^{n} B(x_i)$$

and, according to the factorization test, $T(\mathbf{X})$ is a sufficient statistic.

2.57. The sufficiency of T_n follows from Problems 2.56 and 2.48. We now test it for completeness, i.e., show that only the function φ for which $\varphi(l) = 0$, $l = 0, 1, \ldots, rn$, meets the condition $\mathbf{E}_\theta\varphi(T_n) = 0\ \forall\theta \in (0, 1)$. Due to the reproducibility, $\mathscr{L}_\theta(T_n) = Bi(kn, \theta)$. Consequently, we have

$$\mathbf{E}_\theta\varphi(T_n) = \sum_{l=0}^{kn} \varphi(l)C_{kn}^l\theta^l(1-\theta)^{kn-l},$$

and the completeness condition is equivalent to

$$\sum_{l=0}^{kn} \varphi(l)C_{kn}^l x^l = 0\ \forall x > 0,\ x = \frac{\theta}{1-\theta}.$$

Since the polynomial is identically zero, all its coefficients are zero, which we wanted to prove. The statistic T_n being complete, we conclude that the unbiased

estimators only exist for the parametric functions $\tau(\theta)$ of the form $\mathbf{E}_\theta H(T_n)$, i.e., are the polynomials in θ of degree $\leqslant kn$. Specifically, the statistic $(T_n)_j/(kn)_j$ is an unbiased (and hence optimum) estimator of the degree θ^j for $j \leqslant kn$ (see Problem 1.52), and the linearity of the optimality property implies that the unbiased estimator for the polynomial $\tau(\theta) = \sum_{j=0}^{r} a_j\theta^j$ at $r \leqslant kn$ has the form given in the statement of the problem. This generalizes the results of Problems 2.5, 2.7, and 2.8 making them stronger.

2.58. The sufficiency of the statistic T_n follows from the results obtained in Problems 2.56 and 2.48. We have $\mathscr{L}_\theta(T_n) = \Pi(n\theta)$, and therefore the completeness condition is equivalent to $\sum_{k=0}^{\infty} \varphi(k)\dfrac{(n\theta)^k}{k!} = 0\ \forall\theta > 0$. But since the power series is identically zero, all its coefficients are zero. The completeness condition is only met by the function φ for which $\varphi(k) = 0$, $k = 0, 1, 2, \ldots$. This means that the statistic T_n is complete. It follows from Problem 2.9 that the statistic $(T_n)_j/n^j$ is an optimum unbiased estimator for θ^j. Taking into account that the optimality property is linear, we get the last assertion of the problem.

2.59. We find from the previous problem that the function $\tau(\theta) = \sum_{j=0}^{\infty} (\theta(z-1))^j/j!$ has an optimum estimator of the form

$$\tau^* = \sum_{j=0}^{\infty} \frac{(T_n)_j(z-1)^j}{j!n^j} = \sum_{j=0}^{T_n} C_{T_n}^j \frac{(z-1)^j}{n^j} = \left(1 + \frac{z-1}{n}\right)^{T_n},$$

for the function $\pi_k(\theta) = \sum_{j=0}^{\infty} (-1)^j \dfrac{\theta^{k+j}}{k!j!}$ the optimum estimator is

$$\pi_k^* = \sum_{j=0}^{\infty} (-1)^j \frac{(T_n)_{k+j}}{k!j!n^{k+j}} = \frac{(T_n)_k}{k!n^k} \sum_{j=0}^{T_n-k} (-1)^j \frac{(T_n-k)_j}{n^j j!}$$

$$= C_{T_n}^k \left(1 - \frac{1}{n}\right)^{T_n-k} \frac{1}{n^k},$$

for the function $\tau_r(\theta) = \sum_{k=r}^{\infty} \pi_k(\theta)$ it is

$$\tau_r^* = \sum_{k=r}^{\infty} \pi_k^* = \left(\sum_{k=r}^{T_n} C_{T_n}^k \left(1 - \frac{1}{n}\right)^{T_n-k} \frac{1}{n^k}\right) I(T_n \geqslant r).$$

2.60. (1) We have $f(\mathbf{x}; \theta) = \exp\{A(\theta)B(x) + C(\theta) + D(x)\}$ for $A(\theta) = \ln \theta$, $B(x) = x$, $C(\theta) = -\ln f(\theta)$, $D(x) = \ln a(x)$. The required assertion directly follows from the result of Problem 2.46.

(2) The likelihood function

$$L(\mathbf{x}; \theta) = \theta^{T_n(\mathbf{x})} f^{-n}(\theta) \prod_{i=1}^{n} a(x_i)$$

and the factorization test prove the sufficiency of T_n. To find the distribution of T_n, we recall that its generating function (see the hint) is

$$\mathbf{E}_\theta z^{T_n} = \varphi^n(z; \theta) = f^n(z\theta)/f^n(\theta).$$

Extracting from the right-hand side the coefficient at z^t, we obtain the required result.

Now let $\varphi(t)$ be an arbitrary function defined on the set $\{nl, nl + 1, \ldots\}$ and such that $\mathbf{E}_\theta \varphi(T_n) = 0 \ \forall \theta \in \Theta$, i.e.,

$$\sum_{t=nl}^{\infty} \varphi(t) b_n(t) \theta^t = 0 \ \forall \theta \in \Theta.$$

It follows that $\varphi(t) = 0$ for all t with $b_n(t) \neq 0$, i.e., $\varphi(t) = 0$ on the set of all possible values of the statistic T_n. Thus, T_n is a complete sufficient statistic.

(3) By (2) it is sufficient to verify that $\mathbf{E}_\theta \tau_s^* = \theta^s$. We have

$$\mathbf{E}_\theta \tau_s^* = \sum_{t=nl+s}^{\infty} \frac{b_n(t-s)}{b_n(t)} \mathbf{P}_\theta(T_n = t) = \sum_{t=nl+s}^{\infty} b_n(t-s)\theta^t/f^n(\theta)$$

$$= \theta^s \sum_{t=nl}^{\infty} b_n(t)\theta^t/f^n(\theta) = \theta^s$$

because $\sum_{t=nl}^{\infty} b_n(t)\theta^t = f^n(\theta)$.

The optimality of the estimator $\tau^*(s)$ is provided in a similar way.

(4) Since the optimality property is linear, we use (3) to find

$$\tau^* = \sum_{j=r}^{\infty} a_j \tau_j^* = b_n^{-1}(T_n) \sum_{j=r}^{T_n - nl} a_j b_n(T_n - j) I(T_n \geqslant nl + r).$$

Finally, if $T_n \geqslant (n+1)l$, then we use the hint to find the estimator for f^*.

2.61. This model is a special case of the model from Problem 2.60 (for $f(\theta) = e^\theta - 1$). We therefore have $\tau^* = b_n(T-1)/b_n(T)$, where

$$b_n(k) = \operatorname{coef}_{\theta^k}(e^\theta - 1)^n = \sum_{r=0}^{n} (-1)^{n-r} C_n^r r^k/k! = \Delta^n \theta^k/k!$$

for $k \geqslant n$ and $b_n(k) = 0$ for $k < n$. Whence follows the required result.

2.62. Suppose that in Problem 2.60 $f(\theta) = (1 - \theta)^{-r}$. Taking into account the expansion $f^n(\theta) = (1 - \theta)^{-rn} = \sum_{t=0}^{\infty} C^t_{rn+t-1}\theta^t$ (see Problem 2.11), we find

$$\tau_1^* = C^{T_n-s}_{rn+T_n-s-1}/C^{T_n}_{rn+T_n-1} = \frac{(T_n)_s}{(T_n + rn - 1)_s}$$

(this is a stronger result compared to Problem 2.11).

Since $\tau_2(\theta) = f^{-1}(\theta)$, we find that

$$\tau_2^* = b_{n-1}(T_n)/b_n(T_n) = C^{T_n}_{r(n-1)+T_n-1}/C^{T_n}_{rn+T_n-1} = \frac{(rn - 1)_r}{(T_n + rn - 1)_r}$$

as in Problem 2.60.

We finally obtain

$$\tau_3^* = C^{T_n-s}_{rn+T_n-j-s-1}/C^{T_n}_{rn+T_n-1} = \frac{(T_n)_s}{(T_n + rn - 1)_s}\,\frac{(rn - 1)_j}{(T_n + rn - s - 1)_j}.$$

2.63. The hint to Problem 2.45 implies that the function $f(x; \theta)$ can be represented as

$$f(x; \theta) = \exp\left\{\sum_{j=1}^{N-1} \theta_j'\delta(x, a_j) + \ln p_N\right\},$$

where $\theta_j' = \ln \dfrac{p_j}{p_N}$, $j = 1, \ldots, N - 1$.

By the test for an r-parametric exponential family (for $r = N - 1$) it follows that $\mathbf{T} = (T_1, \ldots, T_{N-1})$, where $T_j = \sum_{i=1}^{n} \delta(X_i, a_j) = \nu_j$, $j = 1, \ldots, N - 1$, is a minimal complete sufficient statistic. Consequently, in the given model the unbiased estimators only exist for the parametric functions of the form $\mathbf{E}_\theta H(\mathbf{T})$. The class of these functions coincides with the class of the polynomials in $p_1, \ldots, p_N$ of degree $\leqslant n$ (see the solution to Problem 2.29). It follows from Problem 1.52 (b) that the statistic $\tau^* = (\nu_1)_{k_1}\ldots(\nu_N)_{k_N}/(n)_{k_1+\ldots+k_N}$ is an optimum estimator for $\tau(\theta) = p_1^{k_1}\ldots p_N^{k_N}$ for $k_1 + \ldots + k_N \leqslant n$. The estimators for arbitrary polynomials are constructed from the linear combinations of these statistics (because the optimality property is linear).

2.64. The model $\mathcal{N}(\theta, \sigma^2)$ is of an exponential type, and its sample mean $\overline{X}$ is its complete sufficient statistic. Therefore, T^* is an optimum unbiased estimator for the function $\tau(\theta) = \theta^2$ (compare to Problem 2.50). The optimality of the estimators in Problem 2.16 is proved in a similar way because T^2 is a complete sufficient statistic for the model $\mathcal{N}(\mu, \theta^2)$.

2.65. (1) We have $\mathbf{E}_\theta T_1 = \mathbf{P}_\theta(\xi \leqslant x_0) = \tau(\theta)$, i.e., T_1 is an unbiased estimator for $\tau(\theta)$. Since $\overline{X}$ is here a complete sufficient statistic (see the solution to Problem 2.64), the optimum estimator can be computed by the formula

$$\tau^* = \mathbf{E}_\theta(T_1|\overline{X}) = \mathbf{P}_\theta(X_1 - \overline{X} \leqslant x_0 - \overline{X}|\overline{X}).$$

But $X_1 - \overline{X}$ and $\overline{X}$ are independent (see Problem 1.56) and $\mathscr{L}_\theta(X_1 - \overline{X}) = \mathscr{N}\left(0, \frac{n-1}{n}\sigma^2\right)$. Therefore,

$$\tau^* = \mathbf{P}_\theta(X_1 - \overline{X} \leqslant x_0 - \overline{X}) = \Phi\left(\sqrt{\frac{n}{n-1}}\,\frac{x_0 - \overline{X}}{\sigma}\right).$$

(2) In the first case we have $\tau_1(\theta; \sigma^2) = \mathbf{E}_\theta\varphi_1(\xi) = \exp\{it\theta - (1/2)\sigma^2t^2\}$, whence $\tau_1^* = \exp\{it\overline{X} - (1/2)\sigma_1^2t^2\}$. In the second case we have $\tau_2(\theta; \sigma^2) = \theta^2 + \sigma^2$, whence $\tau_2^* = \overline{X^2} + \sigma_1^2$ (compare with Problem 2.64). Finally, we have $\tau_3(\theta; \sigma^2) = \Phi\left(\frac{x_0 - \theta}{\sigma}\right)$, and the estimator for τ_3^* coincides with that obtained in (1).

2.66. Here the function $f(x; \theta)$ can be written as

$$f(x; \theta) = \exp\{\theta_1'x + \theta_2'x^2 + c(\theta_1', \theta_2')\},\ \theta_1' = \frac{\theta_1}{\theta_2^2},\ \theta_2' = -\frac{1}{2\theta_2^2}.$$

By the test for an r-parametric exponential family it follows that $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is a minimal complete sufficient statistic. The equivalent pair $(\overline{X}, S^2)$ is also a minimal complete sufficient statistic since the two statistics are in a one-to-one correspondence. This means that the estimators in Problem 2.20 are optimal.

2.67. Since $\mathbf{E}_\theta(S^2) = \frac{n-1}{n}\mu_2 = \frac{n-1}{n}\gamma^2\theta^2$ (see Problem 1.27) and $\mathbf{E}_\theta(\overline{X^2}) = \theta^2 + \frac{1}{n}\gamma^2\theta^2 = \frac{n+\gamma^2}{n}\theta^2$ (see Problem 2.13), we have $\mathbf{E}_\theta\varphi(T) = 0$ for all θ. This means that the test for completeness is not fulfilled.

2.68. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the respective measurements. Then $\mathbf{X}$ is a sample from the distribution $\mathscr{N}(\theta_1, \theta_2^2)$, and we are estimating the parametric function $\tau(\theta) = \frac{\pi}{4}\theta_1^2$. Using the solution to Problem 2.67, we find

$$\mathbf{E}_\theta\left(\overline{X}^2 - \frac{1}{n-1}S^2\right) = \theta_1^2 + \frac{1}{n}\theta_2^2 - \frac{1}{n-1}\,\frac{n-1}{n}\theta_2^2 = \theta_1^2.$$

Since the sufficient statistic $T = (\overline{X}, S^2)$ is complete (see Problem 2.66), we see that the statistic $\tau^* = \frac{\pi}{4}\left(\overline{X}^2 - \frac{1}{n-1}S^2\right)$ is an optimum unbiased estimator for $\tau(\theta)$.

2.69. As stated, the conditional $\varphi_A(T) = \mathbf{P}_\theta(T_1 \in A|T)$ and unconditional $\gamma_A = \mathbf{P}_\theta(T_1 \in A)$ probabilities of the event A do not depend on the parameter θ. We also have $\mathbf{E}_\theta\varphi_A(T) = \gamma_A$, i.e., $\mathbf{E}_\theta g(T) = 0\ \forall\theta$, where $g(T) = \varphi_A(T) - \gamma_A$. This and the completeness of the statistic T give $\varphi_A(T) \equiv \gamma_A$, which means

that the conditional and unconditional probabilities coincide, and the statistics T_1 and T are independent.

2.70. Here $T^2 = X_1^2 + X_2^2 + X_3^2$ is a complete sufficient statistic (see the solution to Problem 2.64), and $T_1 = I(X_1 \leqslant x_0)$ is an unbiased estimator for $\tau(\theta)$. Therefore, the optimum estimator is

$$\tau^* = \mathbf{E}_\theta(T_1|T) = \mathbf{P}_\theta\left(\frac{X_1}{T} \leqslant \frac{x_0}{T}\middle| T\right).$$

The statistic $\eta = X_1/T = Y_1/\sqrt{Y_1^2 + \chi_2^2}$, where $Y_i = X_i/\theta$, $i = 1, 2, 3$, $\chi_2^2 = Y_2^2 + Y_3^2$, has a distribution independent of the parameter θ (because $\mathscr{L}_\theta(Y_i) = \mathscr{N}(0, 1)$, $i = 1, 2, 3$) and, according to the Basu theorem (see Problem 2.69), η and T are independent. We thus have $\tau^* = F_\eta\left(\frac{x_0}{T}\right)$, where $F_\eta(u) = \mathbf{P}(\eta \leqslant u)$. The distribution function $F_\eta(u)$ was found in Problem 1.58 for an arbitrary sample size. We take $n = 4$ to obtain

$$F_\eta(u) = 1 - \frac{1}{2}B\left(1 - u^2; 1, \frac{1}{2}\right) = \frac{1 + u}{2}$$

for $0 \leqslant u \leqslant 1$ and

$$F_\eta(u) = 1 - F_\eta(-u) = \frac{1 + u}{2}$$

for $-1 \leqslant u \leqslant 0$. We thus find $\mathscr{L}(\eta) = R(-1, 1)$ and

$$\tau^* = \begin{cases} 1 & \text{for } x_0 > T, \\ \frac{1}{2}\left(1 + \frac{x_0}{T}\right) & \text{for } |x_0| \leqslant T, \\ 0 & \text{for } x_0 < -T. \end{cases}$$

2.71. We introduce the random variables $Y_i = \frac{X_i - \theta_1}{\theta_2}$, $i = 1, \ldots, n$, whose distributions are independent of $\theta = (\theta_1, \theta_2)$. Then

$$\frac{X_i - \overline{X}}{S(\mathbf{X})} = \frac{Y_i - \overline{Y}}{S(\mathbf{Y})}, \quad i = 1, \ldots, n,$$

i.e., the distribution of the statistic U is independent of θ. Since $T = (\overline{X}, S^2(\mathbf{X}))$ is a complete sufficient statistic for the model $\mathscr{N}(\theta_1, \theta_2^2)$ (see Problem 2.66), the statistics T and U are independent by the Basu theorem (see Problem 2.69).

2.72. Since $\mathbf{E}_\theta T_1 = \mathbf{P}_\theta(X_1 \leqslant x_0) = \tau(\theta)$, T_1 is an unbiased estimator for $\tau(\theta)$. Consequently, the optimum estimator can be found from (in what follows $T = (\overline{X}, S^2)$)

$$\tau^* = \mathbf{E}_\theta(T_1|T) = \mathbf{P}_\theta(X_1 \leqslant x_0|T) = \mathbf{P}_\theta(\eta \leqslant u_0|T),$$

where $\eta = \dfrac{X_1 - \overline{X}}{\sqrt{n-1}\,S}$, $u_0 = \dfrac{x_0 - \overline{X}}{\sqrt{n-1}\,S}$. According to the solution to the previ-

ous problem, the statistics η and T are independent and, therefore, $\tau^* = F_\eta(u_0)$. The distribution function of the statistic η was calculated in Problem 1.58. We use this result to obtain

$$\tau^* = \begin{cases} 1 - \frac{1}{2}\mathrm{B}\left(1 - u_0^2; \frac{n-2}{2}, \frac{1}{2}\right) \text{ for } \overline{X} < x_0, \\ \frac{1}{2}\mathrm{B}\left(1 - u_0^2; \frac{n-2}{2}, \frac{1}{2}\right) \quad \text{for } \overline{X} \geqslant x_0. \end{cases}$$

Published tables for the beta distribution function $\mathrm{B}(t; a, b)$ will help in the calculations.

2.73. The model $\Gamma(\theta, \lambda)$ is of the exponential type, and $T = \sum_{i=1}^{n} X_i$ is its complete sufficient statistic. Then the estimators in Problem 2.21 are optimal. In order to prove the second assertion, it is sufficient to show that there is no function $H(T)$ which meets the condition $\mathrm{E}_\theta H(T) = \theta^{-a}\ \forall\theta > 0$. This condition can be written as

$$\int_0^\infty H_1(x)e^{-zx}\,dx = z^{a-\lambda n} \quad \forall z > 0,$$

where $z = 1/\theta$, $H_1(x) = H(x)x^{\lambda n - 1}/\Gamma(\lambda n)$.

If $m = a - \lambda n + 1$ is an integer, then we differentiate the identity m times with respect to z to obtain

$$\int_0^\infty H_1(x)x^m e^{-zx}\,dx \equiv 0.$$

The integral is the Laplace transform of $x^m H_1(x)$, and therefore for $x > 0$ we have $H(x) = 0$.

2.74. Since T is a complete sufficient statistic, it is enough to show that the estimator τ^* is unbiased. We have $\mathscr{L}_\theta(T) = \Gamma(\theta, \lambda n)$. Consequently,

$$\begin{aligned} \mathrm{E}_\theta\varphi(tT) &= \int_0^\infty \varphi(tx)x^{\lambda n - 1}e^{-x/\theta}\,dx/(\Gamma(\lambda n)\theta^{\lambda n}) \\ &= \int_0^\infty \varphi(y)y^{\lambda n - 1}e^{-y/\theta t}\,dy/(\Gamma(\lambda n)(\theta t)^{\lambda n}) \end{aligned}$$

and

$$\int_0^1 e^{-y/\theta t}\left(\frac{1}{t} - 1\right)^{\lambda(n-1)-1}\frac{dt}{t^2} = e^{-y/\theta}\left(\frac{\theta}{y}\right)^{\lambda(n-1)}\Gamma(\lambda(n-1)).$$

We change the order of integration and get

$$\mathrm{E}_\theta\tau^* = \frac{1}{\Gamma(\lambda)\theta^\lambda}\int_0^\infty \varphi(y)y^{\lambda-1}e^{-y/\theta}\,dy = \mathrm{E}_\theta\varphi(\xi).$$

Specifically, if $a > -\lambda$, then the mean $\mathbf{E}_\theta\varphi(\xi)$ exists for all $\varphi(x) = x^a$ and we have

$$\tau(\theta) = \mathbf{E}_\theta\varphi(\xi) = \frac{\Gamma(\lambda + a)}{\Gamma(\lambda)}\,\theta^a.$$

Therefore,

$$\tau^* = \frac{\Gamma(\lambda n)T^a}{\Gamma(\lambda)\Gamma(\lambda(n-1))}\int_0^1 t^{\lambda+a-1}(1-t)^{\lambda(n-1)-1}\,dt = \frac{\Gamma(\lambda n)\Gamma(a+\lambda)}{\Gamma(\lambda)\Gamma(a+\lambda n)}\,T^a.$$

For $a = 1, 2$ we arrive at the results of Problems 2.48 and 2.51, respectively.

We find a more general result in a similar way, i.e., if $\tau(\theta) = \theta^a e^{-x/\theta}$, $a > -\lambda n$, $x \geqslant 0$, then its optimum estimator is

$$\tau^* = \frac{\Gamma(\lambda n)}{\Gamma(\lambda n + a)}\,\frac{(T-x)^{\lambda n + a - 1}}{T^{\lambda n - 1}}\,I(T \geqslant x).$$

2.75. We have $\tau(\theta; t) = \mathbf{E}_\theta I(\xi \geqslant t) = \mathbf{P}_\theta(\xi \geqslant t)$. Consequently,

$$\tau^* = \frac{\Gamma(\lambda n)}{\Gamma(\lambda)\Gamma(\lambda(n-1))}\int_0^1 I(xT \geqslant t)x^{\lambda-1}(1-x)^{\lambda(n-1)-1}\,dx.$$

Here $I(xT \geqslant t) = 1 \Leftrightarrow x \geqslant t/T$. If $T \leqslant t$, then the integral is equal to zero, and if $T > t$, then it is expressed in terms of the beta distribution function as

$$\int_{t/T}^1 x^{\lambda-1}(1-x)^{\lambda(n-1)-1}\,dx = \mathrm{B}(\lambda, \lambda(n-1))\left(1 - \mathrm{B}\left(\frac{t}{T}; \lambda, \lambda(n-1)\right)\right).$$

Specifically, for $\lambda = 1$ we have

$$\mathrm{B}(x; 1, n-1) = \frac{1}{\mathrm{B}(1, n-1)}\int_0^x (1-z)^{n-2}\,dz = 1 - (1-x)^{n-1},$$

which gives the required result for the exponential distribution.

2.76. We write the likelihood function in the form

$$L(\mathbf{x}; \theta) = \prod_{i=1}^n \frac{\lambda}{\theta^\lambda} x_i^{\lambda-1} e^{-(x_i/\theta)^\lambda} = \frac{\lambda^n}{\theta^{\lambda n}}\, e^{-T(\mathbf{x})/\theta^\lambda} \prod_{i=1}^n x_i^{\lambda-1}.$$

According to the factorization test, T is a sufficient statistic. We look for its distribution. Noting that

$$\mathbf{P}_\theta(2(\xi/\theta)^\lambda \leqslant x) = \mathbf{P}_\theta\left(\xi \leqslant \theta\left(\frac{x}{2}\right)^{1/\lambda}\right) = 1 - e^{-x/2},$$

i.e., $\mathscr{L}_\theta(2(\xi/\theta)^\lambda) = \chi^2(2)$, we find $\mathscr{L}_\theta(2T/\theta^\lambda) = \chi^2(2n)$. The distribution density of T has the form

$$f_T(x) = \frac{x^{n-1}}{\Gamma(n)\theta^{\lambda n}} e^{-x/\theta^\lambda},\ x \geqslant 0.$$

The condition $\mathbf{E}_\theta\varphi(T) = 0\ \forall\theta > 0$ has the form

$$\int_0^\infty \varphi(x)x^{n-1}e^{-zx}\,dx = 0\ \forall z > 0.$$

It follows that $\varphi(x) = 0$, $x > 0$, i.e., the statistic T is complete. It is now sufficient to verify that $\mathbf{E}_\theta\tau^* = \tau(\theta)$. We change the order of integration to obtain

$$\mathbf{E}_\theta\tau^* = (n-1)\int_0^1 (1-t)^{n-2}\left[\int_0^\infty \varphi((tx)^{1/\lambda})f_T(x)\,dx\right]dt$$

$$= \frac{\lambda(n-1)}{\Gamma(n)\theta^{\lambda n}}\int_0^\infty \varphi(y)y^{\lambda n-1}\left[\int_0^1 e^{-\frac{y^\lambda}{\theta^\lambda t}}\left(\frac{1}{t}-1\right)^{n-2}\frac{dt}{t^2}\right]dy$$

$$= \frac{\lambda}{\theta^\lambda}\int_0^\infty \varphi(y)y^{\lambda-1}e^{-(y/\theta)^\lambda}\,dy = \mathbf{E}_\theta\varphi(\xi) = \tau(\theta).$$

For $\varphi(x) = x^\lambda$ we have $\tau(\theta) = \theta^\lambda$ and $\tau^* = (n-1)T\int_0^1 t(1-t)^{n-2}\,dt = T/n$.

2.77. By using the indicators, we write the likelihood function as

$$L(\mathbf{x};\ \theta) = \frac{1}{\theta_2^n}\prod_{i=1}^n I(x_i \geqslant \theta_1)\exp\left\{-\frac{x_i-\theta_1}{\theta_2}\right\}$$

$$= \frac{1}{\theta_2^n}I(x_{(1)} \geqslant \theta_1)\exp\left\{-\frac{n}{\theta_2}(\bar{x}-\theta_1)\right\}.$$

The required assertion follows from the factorization test. We find the estimates $\theta_1^* = (nX_{(1)} - \overline{X})/(n-1)$, $\theta_2^* = (n/(n-1))(\overline{X} - X_{(1)})$ with $\mathbf{D}_\theta\theta_1^* = \theta_2^2/(n(n-1))$, $\mathbf{D}_\theta\theta_2^* = \theta_2^2/(n-1)$.

2.78. Similarly to Problem 2.77, we write the likelihood function as

$$L(\mathbf{x};\ \theta) = \prod_{i=1}^n f(x_i;\ \theta)I(x_{(1)} \geqslant a(\theta)).$$

We see that $X_{(1)}$ is the only one-dimensional sufficient statistic which exists if only $f(x_i;\ \theta)$ can be expanded into the product of two functions one of which depends on x_i and the other depends on θ, i.e., if $f(x;\ \theta) = g(x)/h(\theta)$.

2.79. The solution to the previous problem shows that $X_{(n)}$ is a sufficient statistic. In order to verify its completeness, we must find the distribution of $X_{(n)}$. We have

$$\mathbf{P}_\theta(X_{(n)} \leqslant x) = \mathbf{P}_\theta^n(X_1 \leqslant x) = \left(\frac{x}{\theta}\right)^n, \ 0 \leqslant x \leqslant \theta.$$

If

$$\mathbf{E}_\theta \varphi(X_{(n)}) = \frac{n}{\theta^n} \int\limits_0^\theta \varphi(x) x^{n-1}\, dx = 0 \ \forall \theta > 0,$$

then we differentiate the identity $\int\limits_0^\theta \varphi(x) x^{n-1}\, dx \equiv 0$ with respect to θ and obtain $\varphi(\theta) = 0$, $\theta > 0$. This means that $X_{(n)}$ is complete. Since T^* is an unbiased estimator for θ (see Problem 2.24), it is a function of the complete sufficient statistic and hence is the optimum estimator among all the unbiased estimators. We also find

$$\begin{aligned}\mathbf{E}_\theta(T_\lambda - \theta)^2 &= \mathbf{E}_\theta[\lambda(T^* - \theta) + (\lambda - 1)\theta]^2 \\ &= \lambda^2 \mathbf{D}_\theta T^* + (\lambda - 1)^2\theta^2 = \psi(\lambda)\theta^2,\end{aligned}$$

where (see the solution to Problem 2.24) $\psi(\lambda) = \dfrac{\lambda^2}{n(n+2)} + (\lambda - 1)^2_*$. We have $\min\limits_\lambda \psi(\lambda) = \psi(\lambda^*) = \dfrac{1}{(n+1)^2}$ and $\lambda^* = \dfrac{n(n+2)}{(n+1)^2}$. Thus,

$$\frac{\theta^2}{(n+1)^2} < \frac{\theta^2}{n(n+2)} = \mathbf{D}_\theta T^*$$

is a mean square error for the estimator $T_{\lambda^*} = \dfrac{n+2}{n+1} X_{(n)}$.

We prove the optimality of τ^* by verifying the relation $\mathbf{E}_\theta \tau^* = \tau(\theta)\ \forall \theta > 0$.

2.80. For the model $R(\theta_1, \theta_2)$ the likelihood function can be written as $L(\mathbf{x}; \theta) = I(\theta_2 \geqslant x_{(n)}) I(x_{(1)} \geqslant \theta_1)/(\theta_2 - \theta_1)^n$. Consequently, $\mathbf{T} = (X_{(1)}, X_{(n)})$ is a sufficient statistic whose distribution density $\mathbf{T}$ was given in Problem 1.36. We use this result to find that the condition $\mathbf{E}_\theta \varphi(\mathbf{T}) = 0\ \forall \theta$ is equivalent to

$$\int\limits_{\theta_1}^{\theta_2} \int\limits_{x_1}^{\theta_2} \varphi(x_1, x_2)(x_2 - x_1)^{n-2}\, dx_2\, dx_1 = 0 \ \forall \theta.$$

Differentiating this identity first with respect to θ_1 and then with respect to θ_2, we reduce it to the identity $\varphi(\theta_1, \theta_2)(\theta_2 - \theta_1)^{n-2} = 0\ \forall \theta$. It follows that $\varphi(\theta_1, \theta_2) = 0\ \forall \theta_1 < \theta_2$, i.e., the statistic $\mathbf{T}$ is complete. The estimators constructed in Problem 2.25 are optimal as the functions of $\mathbf{T}$. Finally, since $\mathbf{E}_\theta X_{(1)} = \dfrac{n\theta_1 + \theta_2}{n+1}$, $\mathbf{E}_\theta X_{(n)} = \dfrac{\theta_1 + n\theta_2}{n+1}$ (see Problem 1.36), the statistics

$(nX_{(1)} - X_{(n)})/(n-1)$ and $(nX_{(n)} - X_{(1)})/(n-1)$ are unbiased, and hence optimal, estimators for the parameters θ_1 and θ_2, respectively.

2.81. Like in the previous problem, we write the likelihood function as $L(\mathbf{x}; \theta) = I(x_{(1)} \geqslant a(\theta))I(b(\theta) \geqslant x_{(n)})/(b(\theta) - a(\theta))^n$ and see that the statistic $\mathbf{T}$ is sufficient. If we have $a(\theta)\uparrow$, $b(\theta)\downarrow$ as θ grows, then

$$\{x_{(1)} \geqslant a(\theta),\ x_{(n)} \leqslant b(\theta)\} \Leftrightarrow \{\theta \leqslant a^{-1}(x_{(1)}),\ \theta \leqslant b^{-1}(x_{(n)})\}$$
$$\Leftrightarrow \{\theta \leqslant T_1(\mathbf{x}) = \min(a^{-1}(x_{(1)}),\ b^{-1}(x_{(n)}))\},$$

and the likelihood function can be written in the form $L(\mathbf{x}; \theta) = I(T_1(\mathbf{x}) \geqslant \theta)/(b(\theta) - a(\theta))^n$. This means that in our case a one-dimensional sufficient statistic exists and has the form $T_1(\mathbf{X}) = \min(a^{-1}(X_{(1)}), b^{-1}(X_{(n)}))$. Similarly, if we have $a(\theta)\downarrow$, $b(\theta)\uparrow$ as θ grows, a one-dimensional statistic exists and has the form $T_2(\mathbf{X}) = \max(a^{-1}(X_{(1)}), b^{-1}(X_{(n)}))$. There are no other one-dimensional sufficient statistics in the model $R(a(\theta), b(\theta))$.

For the model $R(-\theta, \theta)$ the statistic is

$$T_2(\mathbf{X}) = \max(-X_{(1)}, X_{(n)}) = \max(|X_{(1)}|, |X_{(n)}|).$$

2.82. The condition $\mathbf{E}_\theta \varphi(X) = 0\ \forall\theta$ can be written as

$$\varphi(-1)\frac{\theta}{(1-\theta)^2} + \sum_{x=0}^{\infty} \varphi(x)\theta^x = 0$$

or (see the hint to Problem 2.11) as

$$\varphi(0) + \sum_{x=1}^{\infty} [\varphi(x) + x\varphi(-1)]\theta^x = 0.$$

The functions φ for which $\varphi(0) = 0$, $\varphi(x) = -x\varphi(-1)$, $x = 1, 2, \ldots$, meet this condition. The only bounded function of this type is $\varphi(x) = 0$, $x = -1, 0, 1, \ldots$. Thus, X is a boundedly complete sufficient statistic which is not complete.

2.83. The distribution of the data $\mu = (\mu_1, \ldots, \mu_n)$ is in the set $\mathscr{L}_{N,n}$ of the vectors $\mathbf{l} = (l_1, \ldots, l_n)$ with integer non-negative components, which satisfy the conditions

$$l_1 + \ldots + l_n \leqslant N,\ l_1 + 2l_2 + \ldots + nl_n = n.$$

We find the likelihood function for $\mathbf{P}_N(\mu = \mathbf{l})$, $\mathbf{l} \in \mathscr{L}_{Nn}$. The total number of possible outcomes of the experiment is N^n, and the number of the outcomes compatible with the event $\{\mu = \mathbf{l}\}$ can be calculated in the following way. We first register the elements $k = l_1 + \ldots + l_n$ of the population U, which are in the sample. This can be done in C_N^k different ways. For every such subset the number of ways to form a sample from the elements of the given subset with the given value $\mathbf{l}$ of the statistic μ is the same under the condition that all the k elements must be in the sample. We denote this number $A(\mathbf{l}; k, n)$. Then the total number of favourable outcomes is $C_N^k A(\mathbf{l}; k, n)$ and by the classi-

cal definition of probability we have

$$\mathbf{P}_N(\mu = \mathbf{l}) = g(k;\ N)A(\mathbf{l};\ k,\ n),$$

where the factor $g(k;\ N) = C_N^k N^{-n}$ depends on the parameter N, and only depends on the sampling data through the value of the statistic η which is $k = l_1 + \ldots + l_n$-compatible with the event $\{\mu = \mathbf{l}\}$, while the factor $A(\mathbf{l};\ k,\ n)$ is independent of the parameter N. By the factorization test η is a sufficient statistic for N. To prove the completeness of η, we must show that $\varphi(k) = 0$ for all possible values of the statistic η ($\forall N \geqslant 1$) for any function $\varphi(k)$ from $\mathbf{E}_N\varphi(\eta) = 0\ \forall N$. The distribution of η is (see the solution to Problem 2.37)

$$\mathbf{P}_N(\eta = k) = g(k;\ N) \sum_{j=0}^{k} (-1)^{k-j} C_k^j j^n,\ k \in \mathcal{N}(N) = \{1, 2, \ldots, \min(n, N)\}.$$

We verify that $\varphi(1) = \varphi(2) = \ldots = \varphi(n) = 0$. If $N = 1$, then

$$\mathbf{E}_1\varphi(\eta) = \varphi(1)g(1;\ 1) = \varphi(1) = 0.$$

If $N = 2$, then

$$\mathbf{E}_2\varphi(\eta) = \varphi(1)\mathbf{P}_2(\eta = 1) + \varphi(2)\mathbf{P}_2(\eta = 2) = \varphi(2)g(2;\ 2)(2^n - 2) = 0,$$

i.e., $\varphi(2) = 0$. Similarly, putting $N = 3$, we find from the condition $\mathbf{E}_3\varphi(\eta) = 0$ and the equations $\varphi(1) = \varphi(2) = 0$ that $\varphi(3) = 0$, etc. We thus see that $\varphi(k) = 0$ for all $k \leqslant n$, and the statistic η is complete.

2.84. The efficiency test implies that

$$a(\theta)\,\frac{\partial \ln L(\mathbf{x};\ \theta)}{\partial\theta} = \tau^* - \tau(\theta),$$

whence follows that $\hat{\theta}_n$ satisfies the equation $\tau(\theta) = \tau^*$. We calculate its second derivative

$$\frac{\partial^2 \ln L(\mathbf{x};\ \theta)}{\partial\theta^2} = -\frac{a(\theta)\tau'(\theta) + (\tau^* - \tau(\theta))a'(\theta)}{a^2(\theta)}$$

to show that it is unique. Since we are dealing with the exponential model (see the solution to Problem 2.46), $\mathbf{D}_\theta\tau^* = a(\theta)\tau'(\theta) > 0$ and, therefore, $\left.\dfrac{\partial^2 \ln L(\mathbf{x};\ \theta)}{\partial\theta^2}\right|_{\theta=\hat{\theta}_n} < 0$, i.e., every solution of $\tau(\theta) = \tau^*$ is a local maximum of the likelihood function. For more than one maximum there would have been a minimum between the maximums $\left(\text{i.e., at the points of minimum which also meet the equation } \tau(\theta) = \tau^* \text{ we would have } \dfrac{\partial^2 \ln L}{\partial\theta^2} > 0\right)$. Since there is no minimum, we conclude that only one maximum may exist. Accordingly, we arrive at the following table for the values of $\hat{\theta}_n$ in some of the models (use Problem 2.48):

Model	$\mathcal{N}(\theta, \sigma^2)$	$\mathcal{N}(\mu, \theta^2)$	$\Gamma(\theta, \lambda)$	$B(\theta, 1)$	$Bi(k, \theta)$	$\Pi(\theta)$	$\overline{Bi}(r, \theta)$
$\hat{\theta}_n$	$\overline{X}$	T_1	$\overline{X}/\lambda$	T_2	$\overline{X}/k$	$\overline{X}$	$\overline{X}/(r + \overline{X})$

Here

$$T_1 = \left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2\right]^{1/2}, \; T_2 = \left[-\frac{1}{n} \sum_{i=1}^{n} \ln X_i\right]^{-1}.$$

2.85. In the distribution $\mathcal{N}(\theta, \sigma^2)$ the point θ is a theoretical median, and therefore the sample median T_n is in the given case a consistent estimator for θ whose distribution, according to the solution of Problem 1.32, satisfies the asymptotic relation $\mathcal{L}_\theta(T_n) \sim \mathcal{N}(\theta, \pi\sigma^2/2n)$, i.e., $\sigma_T^2(\theta) = \pi\sigma^2/2$. Then $\hat{\theta}_n = \overline{X}$ (see the solution to Problem 2.84) and $\mathcal{L}_\theta(\hat{\theta}_n) = \mathcal{N}(\theta, \sigma^2/n)$. From this we get $\text{eff}\,(T_n; \theta) = \sigma^2/\sigma_T^2(\theta) = 2/\pi = 0.637\ldots$. This means that for large n the sample mean $\overline{X}$ for the sample of size $n' = 2n/\pi$ estimates θ with the same accuracy as the sample median $X_{([n/2]+1)}$ of the sample of size n, whatever the values of θ and σ^2.

2.86. We have

$$L(\mathbf{x};\, \theta) = \frac{1}{(\sqrt{2\pi}\,\theta_2)^n} \exp\left\{-\frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \theta_1)^2\right\}$$
$$= \frac{1}{(\sqrt{2\pi}\,\theta_2)^n} \exp\left\{-\frac{n}{2\theta_2^2} (s^2 + (\bar{x} - \theta_1)^2)\right\},$$

where $s^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2$, $\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$. The likelihood equations $\dfrac{\partial \ln L}{\partial \theta_1} = \dfrac{\partial \ln L}{\partial \theta_2} = 0$ become

$$\bar{x} = \theta_1, \; \theta_2^2 = s^2 + (\bar{x} - \theta_1)^2.$$

They uniquely define the solution $\theta = (\bar{x}, s)$. Let us show that this is the point of maximum of the likelihood function. We rewrite $L(\mathbf{x};\, \theta)$ in the form

$$L(\mathbf{x};\, \theta) = (2\pi e s^2)^{-n/2} \exp\{-n\psi(\mathbf{x};\, \theta)\},$$

where $\psi(\mathbf{x};\, \theta) = \dfrac{(\bar{x} - \theta_1)^2}{2\theta_2^2} + \dfrac{1}{2}\left(\dfrac{s^2}{\theta_2^2} - 1\right) - \ln\dfrac{s}{\theta_2}$. We shall minimize the function $\psi(\mathbf{x};\, \theta)$ in $\theta = (\theta_1, \theta_2)$. The estimate $\ln a \leqslant a - 1 \;\forall a > 0$ is easily found. Putting $a = b^2$, we obtain the estimate $\ln b \leqslant (b^2 - 1)/2 \;\forall b > 0$ (the

equality only holds for $b = 1$). We now have $\psi(\mathbf{x}; \theta) \geq 0$ (the equality only holds at the point $\theta = (\bar{x}, s)$). Consequently, $L(\mathbf{x}; \theta) \leq (2\pi e s^2)^{-n/2}$ (with the equation at the point $\theta = (\bar{x}, s)$). We have thus proved that $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n})$ exists, is unique, and $\hat{\theta}_n = (\overline{X}, S)$.

2.87. We find the form of $\hat{\tau}_n$ from Problem 2.86 and the invariance property of the maximum likelihood estimates. For $n \to \infty$ we have $\mathcal{L}_\theta(\sqrt{n}(\hat{\tau}_n - \tau(\theta))) \to \mathcal{N}(0, \sigma_\tau^2(\theta))$, where (see Problem 2.44)

$$\sigma_\tau^2(\theta) = \mathbf{b}'(\theta)\mathbf{I}^{-1}(\theta)\mathbf{b}(\theta) = \left(\frac{\partial\tau(\theta)}{\partial\theta_1}\right)^2 \theta_2^2 + \left(\frac{\partial\tau(\theta)}{\partial\theta_2}\right)^2 \theta_2^2/2$$

$$= \frac{1}{2\pi} e^{-\frac{(x_0-\theta_1)^2}{\theta_2^2}} \left[1 + \frac{(x_0 - \theta_1)^2}{2\theta_2^2}\right].$$

2.88. The form of $\hat{\theta}_n$ is given in Problem 2.84. Using it, we may write

$$\hat{\theta}_n = \frac{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\Gamma\left(\frac{n}{2}\right)} \tau_1^*,$$

where τ_1^* is the unbiased estimator for θ obtained in Problem 2.16. We find $\mathbf{E}_\theta\hat{\theta}_n = c_n\theta$, $c_n = \dfrac{\sqrt{2}\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\Gamma\left(\frac{n}{2}\right)}$. Stirling's formula for the gamma function $\Gamma(z) \sim \sqrt{2\pi z}z^{z-1}e^{-z}$, $z \to \infty$, gives $c_n \to 1$ as $n \to \infty$, and we see that $\hat{\theta}_n$ is asymptotically unbiased. We then have

$$\mathbf{D}_\theta\hat{\theta}_n = \mathbf{E}_\theta\hat{\theta}_n^2 - (\mathbf{E}_\theta\hat{\theta}_n)^2 = \theta^2(1 - c_n^2) \to 0, \ n \to \infty,$$

whence follows that $\hat{\theta}_n$ is consistent.

From Problem 2.43 we infer that $\mathcal{L}_\theta(\sqrt{n}(\hat{\theta}_n - \theta)) \to \mathcal{N}(0, i^{-1}(\theta)) = \mathcal{N}(0, \theta^2/2)$ as $n \to \infty$, and consider the estimator $T_n = \sqrt{\dfrac{\pi}{2}}\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} |X_i - \mu|$ (see Problem 2.15). Using the Central Limit Theorem and the solution to Problem 2.15, we find that $\mathcal{L}_\theta(T_n) \sim \mathcal{N}\left(\theta, \dfrac{\pi - 2}{2n}\theta^2\right)$ as $n \to \infty$. The asymptotic efficiency is

$$\text{eff}\,(T_n; \theta) = \frac{\theta^2}{2}\bigg/\frac{\pi - 2}{2}\theta^2 = \frac{1}{\pi - 2} = 0.88\ldots$$

2.89. The likelihood function is

$$L(\mathbf{x};\ \theta) = \frac{1}{(4\pi\theta)^{n/2}} \exp\left\{-\frac{1}{4\theta}\sum_{i=1}^{n}(x_i - \theta)^2\right\}.$$

The likelihood equation is

$$\theta^2 + 2\theta - T_n = 0,\ T_n = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

This equation has the only positive solution $\hat{\theta}_n = \sqrt{1 + T_n} - 1$, which maximizes $L(\mathbf{x};\ \theta)$. By the law of large numbers T_n converges in probability to $\mathbf{E}_\theta X_1^2 = \mathbf{D}_\theta X_1 + (\mathbf{E}_\theta X_1)^2 = 2\theta + \theta^2$ as $n \to \infty$. Consequently,

$$\hat{\theta}_n \xrightarrow{\mathbf{P}_\theta} \sqrt{1 + 2\theta + \theta^2} - 1 = \theta.$$

2.90. The distribution density of the observations is

$$f(x,\ y;\ \theta) = \frac{1}{2\pi}\exp\left\{-\frac{x^2 + y^2}{2\sigma^2(1 - \varrho^2)} + \frac{\varrho xy}{\sigma^2(1 - \varrho^2)} - \frac{1}{2}\ln\left[\sigma^4(1 - \varrho^2)\right]\right\}.$$

Taking into account the hint, we may write

$$f(x,\ y;\ \mathbf{q}) = \frac{1}{2\pi}\exp\{q_1(x^2 + y^2) + q_2xy + \tau(\mathbf{q})\},$$

where $\tau(\mathbf{q}) = \dfrac{1}{2}\ln(4q_1^2 - q_2^2)$. The likelihood equations from which we find $\hat{q}_1$ and $\hat{q}_2$ have the form

$$\frac{1}{n}\sum_{i=1}^{n}(x_i^2 + y_i^2) = -\frac{\partial\tau(\mathbf{q})}{\partial q_1} = -\frac{4q_1}{4q_1^2 - q_2^2},$$

$$\frac{1}{n}\sum_{i=1}^{n}x_iy_i = -\frac{\partial\tau(\mathbf{q})}{\partial q_2} = \frac{q_2}{4q_1^2 - q_2^2}.$$

Since $\sigma^2 = -\dfrac{2q_1}{4q_1^2 - q_2^2}$, $\varrho = -\dfrac{q_2}{2q_1}$, the equations directly give the required maximum likelihood estimates

$$\hat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n}(X_i^2 + Y_i^2),\ \hat{\varrho} = 2\sum_{i=1}^{n}X_iY_i\Big/\sum_{i=1}^{n}(X_i^2 + Y_i^2).$$

2.91. The distribution density in our model is

$$f(x, y; \theta) = \frac{1}{2\pi(1-\theta^2)^{1/2}} \exp\left\{-\frac{1}{2(1-\theta^2)}(x^2 - 2\theta xy + y^2)\right\},$$

therefore,

$$\ln L = -n\ln(2\pi) - \frac{n}{2}\ln(1-\theta^2) - \frac{n}{2(1-\theta^2)}(T_{11} - 2\theta T_{12} + T_{22}),$$

where

$$T_{11} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad T_{12} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i, \quad T_{22} = \frac{1}{n}\sum_{i=1}^{n} y_i^2.$$

We thus find

$$\frac{\partial \ln L}{\partial \theta} = \frac{n\theta}{1-\theta^2} - \frac{n\theta}{(1-\theta^2)^2}(T_{11} - 2\theta T_{12} + T_{22}) + \frac{n}{1-\theta^2}T_{12}.$$

The likelihood equation is reduced to the form

$$\theta(1-\theta^2) + (1+\theta^2)T_{12} - \theta(T_{11} + T_{22}) = 0.$$

This cubic equation has three roots, two of which may be complex. If all the three roots are real and belong to the interval $(-1, 1)$, then the one which maximizes the likelihood function L is chosen as $\hat{\theta}_n$. We use the substitution $\theta = x + \frac{1}{3}T_{12}$ to reduce the likelihood equation to the canonical form

$$x^3 + 3p_n x + 2q_n = 0, \quad 3p_n = T_{11} + T_{22} - \frac{1}{3}T_{12}^2 - 1,$$

$$2q_n = T_{12}\left(\frac{1}{3}(T_{11} + T_{22}) - \frac{2}{27}T_{12}^2 - 1\right).$$

If the inequality $p_n^3 + q_n^2 > 0$ holds, the equation has one real root (see the solution to Problem 2.19), which is obvious for $p_n > 0$. Since the sampling moments converge in probability as $n \to \infty$ to the respective theoretical moments (see the solution to Problem 1.38), p_n converges in probability to the quantity $\frac{1}{3}\left(1 + 1 - \frac{1}{3}\theta^2 - 1\right) = \frac{1}{3}\left(1 - \frac{1}{3}\theta^2\right) > 0$. We see that for large n the likelihood equation has one real root which is the maximum likelihood estimate $\hat{\theta}_n$.

The asymptotic variance of $\hat{\theta}_n$ is $1/i_n(\theta)$, where

$$i_n(\theta) = -\mathbf{E}_\theta \frac{\partial^2 \ln L}{\partial\theta^2}$$

$$= -n\mathbf{E}_\theta\left[\frac{1+\theta^2}{(1-\theta^2)^2} + \frac{4\theta}{(1-\theta^2)^2}T_{12} - \frac{1+3\theta^2}{(1-\theta^2)^3}(T_{11} - 2\theta T_{12} + T_{22})\right]$$

$$= -n\left[\frac{1+\theta^2}{(1-\theta^2)^2} + \frac{4\theta^2}{(1-\theta^2)^2} - \frac{1+3\theta^2}{(1-\theta^2)^3}(2-2\theta^2)\right] = n\,\frac{1+\theta^2}{(1-\theta^2)^2}.$$

Thus,

$$\mathscr{L}_\theta(\sqrt{n}(\hat{\theta}_n - \theta)) \to \mathscr{N}\left(0, \frac{(1-\theta^2)^2}{1+\theta^2}\right)$$

as $n \to \infty$.

2.92. By the Central Limit Theorem the statistic T_n is asymptotically normal with centre at θ and the asymptotic variance $\frac{1}{n}\mathbf{D}_\theta(X_1 Y_1) = \frac{1}{n}[\mathbf{E}_\theta(X_1^2 Y_1^2) - \theta^2]$. Therefore, the problem is reduced to calculating the mixed moment $\mathbf{E}_\theta(X_1^2 Y_1^2)$. Here the characteristic function is

$$\varphi(t_1, t_2) = \mathbf{E}_\theta e^{i(t_1X_1 + t_2Y_1)} = \exp\left\{-\frac{1}{2}(t_1^2 + 2\theta t_1 t_2 + t_2^2)\right\}$$

and

$$\mathbf{E}_\theta(X_1^2 Y_1^2) = \left.\frac{\partial^4\varphi(t_1, t_2)}{\partial t_1^2\,\partial t_2^2}\right|_{t_1=t_2=0} = 1 + 2\theta^2.$$

Thus, $\mathbf{D}_\theta(X_1 Y_1) = 1 + \theta^2$, and the asymptotic efficiency of T_n is

$$\operatorname{eff}(T_n;\, \theta) = \left(\frac{1-\theta^2}{1+\theta^2}\right)^2.$$

2.93. (1) Consider the likelihood function

$$L(\mathbf{x};\, \theta) = \prod_{i=1}^{n} f(\mathbf{x}_i;\, \theta)$$

$$= [(2\pi)^k|\Sigma|]^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \mu)'\Sigma^{-1}(\mathbf{x}_i - \mu)\right\}.$$

Here

$$\sum_{i=1}^{n}(\mathbf{x}_i - \mu)'\Sigma^{-1}(\mathbf{x}_i - \mu)$$

$$= \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})'\Sigma^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) + n(\bar{\mathbf{x}} - \mu)'\Sigma^{-1}(\bar{\mathbf{x}} - \mu).$$

Easily verifying the equation $\mathbf{y}'\mathbf{B}\mathbf{y} = \operatorname{tr}(\mathbf{B}\mathbf{Y})$, where $\mathbf{Y} = \mathbf{y}\mathbf{y}'$, and the linearity of the operator tr, we obtain

$$\sum_{l=1}^{n} (\mathbf{x}_l - \overline{\mathbf{x}})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_l - \overline{\mathbf{x}}) = n \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}(\mathbf{x})).$$

These relations give the formula from the hint, which implies that by maximizing the function $L(\mathbf{x};\, \theta)$ in θ, we minimize the function

$$\psi(\mathbf{x};\, \mu,\, \Sigma) = (\overline{\mathbf{x}} - \mu)'\Sigma^{-1}(\overline{\mathbf{x}} - \mu) + [\operatorname{tr}(\Sigma^{-1}\hat{\Sigma}(\mathbf{x})) - k - \ln|\Sigma^{-1}\hat{\Sigma}(\mathbf{x})|]$$

in μ and Σ. We use $\lambda_1, \ldots, \lambda_k$ to denote the roots of the characteristic equation $|\Sigma^{-1}\hat{\Sigma}(\mathbf{x}) - \lambda \mathbf{E}_k| = 0$ or the equivalent equation $|\hat{\Sigma}(\mathbf{x}) - \lambda\Sigma| = 0$. The expression in the brackets is equal to $\lambda_1 + \ldots + \lambda_k - k - \ln(\lambda_1 \ldots \lambda_k)$, and we may write

$$\psi(\mathbf{x};\, \mu,\, \Sigma) = (\overline{\mathbf{x}} - \mu)'\Sigma^{-1}(\overline{\mathbf{x}} - \mu) + \sum_{l=1}^{k} (\lambda_l - 1 - \ln \lambda_l).$$

Since the matrix Σ^{-1} is positive definite and $\lambda - 1 - \ln\lambda \geqslant 0\ \forall\lambda > 0$, the latter relation gives $\psi(\mathbf{x};\, \mu,\, \Sigma) \geqslant 0$ with the equation holding only for $\mu = \overline{\mathbf{x}}$ and $\lambda_1 = \ldots = \lambda_k = 1$, i.e., when $\Sigma = \hat{\Sigma}(\mathbf{x})$.

(2) According to Problem 2.4, the statistic $\dfrac{n}{n-1} S_{ij}$ is an unbiased estimator for σ_{ij}. Consequently, $\mathbf{E}\left(\dfrac{n}{n-1}\hat{\Sigma}\right) = \Sigma$.

(3) To find the required expression for $\max_{\theta} L(\mathbf{x};\, \theta)$, we must take into account that $\operatorname{tr}(\hat{\Sigma}^{-1}(\mathbf{x})\hat{\Sigma}(\mathbf{x})) = \operatorname{tr}\mathbf{E}_k = k$.

2.94. We calculate the moments $\mathbf{E}_\theta X_1^k$ in the following way. Suppose that $\varphi(t) = \mathbf{E}_\theta e^{itY_1} = \exp\left\{it\theta_1 - \dfrac{t^2}{2}\theta_2^2\right\}$. Then

$$\mathbf{E}_\theta X_1^k = \mathbf{E}_\theta e^{kY_1} = \varphi\left(\frac{k}{i}\right) = \exp\left\{k\theta_1 + \frac{k^2}{2}\theta_2^2\right\}.$$

Whence

$$\tau_1(\theta) = \mathbf{E}_\theta X_1 = \exp[\theta_1 + \theta_2^2/2],$$
$$\tau_2(\theta) = \mathbf{D}_\theta X_1 = \mathbf{E}_\theta X_1^2 - (\mathbf{E}_\theta X_1)^2 = \tau_1^2(\theta)(e^{\theta_2^2} - 1).$$

Since $(\hat{\theta}_{1n}, \hat{\theta}_{2n}^2) = (\overline{Y}, S^2(\mathbf{Y}))$, by the invariance property of maximum likelihood estimates we have

$$\hat{\tau}_{1n} = \exp[\overline{Y} + S^2(\mathbf{Y})/2],\quad \hat{\tau}_{2n} = \hat{\tau}_{1n}^2(e^{S^2(\mathbf{Y})} - 1).$$

Since $\overline{Y}$ and $S^2(\mathbf{Y})$ are independent (see Problem 1.56), we have $\mathbf{E}_\theta\hat{\tau}_{1n} = \mathbf{E}_\theta e^{\overline{Y}}\mathbf{E}_\theta e^{S^2(\mathbf{Y})/2}$. Here $\mathscr{L}_\theta(\overline{Y}) = \mathscr{N}(\theta_1,\ \theta_2^2/n)$ and, as before, $\mathbf{E}_\theta e^{\overline{Y}} =$

$\exp\{\theta_1 + \theta_2^2/(2n)\}$. We also have $\mathcal{L}_\theta(nS^2(\mathbf{Y})/\theta_2^2) = \chi^2(n-1)$ (see, for example, the solution to Problem 1.58), whence $\mathbf{E}_\theta e^{S^2(\mathbf{Y})/2} = \psi(\theta_2^2/(2in))$, where $\psi(t) = \mathbf{E}e^{it\chi^2_{n-1}} = (1-2it)^{-(n-1)/2}$ (see the solution to Problem 1.39). From this we find $\mathbf{E}_\theta e^{S^2(\mathbf{Y})/2} = \left(1 - \dfrac{\theta_2^2}{n}\right)^{-(n-1)/2}$ and, finally,

$$\mathbf{E}_\theta \hat{\tau}_{1n} = \tau_1(\theta)\exp\left\{-\frac{n-1}{2n}\theta_2^2\right\}\left(1-\frac{\theta_2^2}{n}\right)^{-(n-1)/2} \to \tau_1(\theta)$$

as $n \to \infty$.

2.96. Here

$$\frac{\partial}{\partial\theta}\ln L(\mathbf{x};\theta) = \frac{\partial}{\partial\theta}\ln\left(\theta^{n\bar{x}}f^{-n}(\theta)\prod_{i=1}^{n}a(x_i)\right)$$

$$= \frac{n\bar{x}}{\theta} - nf'(\theta)/f(\theta) = \frac{n}{\theta}(\bar{x} - \mu(\theta))$$

with $\mu(\theta) = \theta f'(\theta)/f(\theta) = \sum_x xa(x)\theta^x/f(\theta) = \mathbf{E}_\theta\xi$. This means that the likelihood equation has the indicated form. The asymptotic variance $\hat{\theta}_n$ of the estimate is $(ni(\theta))^{-1}$, where the information function is

$$i(\theta) = -\mathbf{E}_\theta\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta) = \mathbf{E}_\theta\left(\frac{\xi-\mu(\theta)}{\theta^2} + \frac{\mu'(\theta)}{\theta}\right) = \frac{\mu'(\theta)}{\theta}.$$

Specifically, in the model $\overline{Bi}(r,\theta)$ the mean is $\mu(\theta) = r\theta/(1-\theta)$, and $\hat{\theta}_n = \overline{X}/(r+\overline{X})$ is the solution to the equation $\mu(\theta) = \overline{X}$ (this coincides with the result of Problem 2.84), while the information function is $i(\theta) = \dfrac{\mu'(\theta)}{\theta} = \dfrac{r}{\theta(1-\theta)^2}$ (the result of Problem 2.43).

2.97. In this case $f(\theta) = e^\theta - 1$, $\mu(\theta) = \theta/(1-e^{-\theta})$, and the likelihood equation $\theta = \bar{x}(1-e^{-\theta})$ cannot be solved exactly. In order to approximately calculate the m.l.e. $\hat{\theta}_n$, we use the accumulation method. Here (see the solution to Problem 2.96)

$$U(\mathbf{x};\theta) = \frac{\partial}{\partial\theta}\ln L(\mathbf{x};\theta) = \frac{n}{\theta}(\bar{x}-\mu(\theta)),\quad i(\theta) = \frac{1-(1+\theta)e^{-\theta}}{\theta(1-e^{-\theta})^2}$$

and the sought-for equations have the form

$$\theta_{k+1} = \theta_k + U(\mathbf{x};\theta_k)/(ni(\theta_k)),\ k = 0, 1, \ldots.$$

2.98. In order to write these equations (see the solution to Problem 2.97), we should know the contribution function $U(\theta) = \dfrac{\partial\ln L(\theta)}{\partial\theta}$ and the information function $i_n(\theta) = -\mathbf{E}_\theta\dfrac{\partial U(\theta)}{\partial\theta}$. In this case

$$L(\theta) = \frac{n!}{h_1! \dots h_N!} \prod_{i=1}^{N} p_i^{\nu_i}(\theta), \; U(\theta) = \sum_{i=1}^{N} \frac{\nu_i}{p_i(\theta)} p_i'(\theta),$$

$$i_n(\theta) = -\sum_{i=1}^{N} \frac{p_i''(\theta)p_i(\theta) - (p_i'(\theta))^2}{p_i^2(\theta)} \mathbf{E}_\theta \nu_i = n \sum_{i=1}^{N} \frac{p_i'(\theta)^2}{p_i(\theta)}$$

because $\mathbf{E}_\theta \nu_i = np_i(\theta)$ and $\sum_{i=1}^{N} p_i(\theta) = 1 \; \forall \theta$.

2.99. Since $i(\theta) = 1/2$ (see Problem 2.43), the iteration consists of

$$\theta_{k+1} = \theta_k + \frac{2}{n} U(\theta_k), \; k = 0, 1, \dots, \; U(\theta) = 2 \sum_{i=1}^{n} \frac{x_i - \theta}{1 + (x_i - \theta)^2}.$$

We may use the sample median $T_n = X_{([n/2]+1)}$ as a first approximation for θ_0, the sample median being a consistent estimator for θ because here the theoretical median coincides with θ. From Problem 1.32 we find that as $n \to \infty$

$$\mathscr{L}_\theta(T_n) \sim \mathscr{N}\left(\theta, \frac{1}{4nf^2(\theta; \theta)}\right) = \mathscr{N}\left(\theta, \frac{\pi^2}{4n}\right).$$

Consequently,

$$\text{eff}\,(T_n; \theta) = \frac{8}{\pi^2} = 0.8\dots.$$

2.100. We write the likelihood function in the form $L(\mathbf{x}; \theta) = I(\theta \geqslant x_{(n)})/\theta^n$ (see the solution to Problem 2.78) and see that it is monotone decreasing in θ for $\theta \geqslant x_{(n)}$, i.e., it attains maximum at $\theta = x_{(n)}$. Then $\hat{\theta}_n = X_{(n)}$. Using the solution to Problem 2.24, we find $\mathbf{E}_\theta \hat{\theta}_n = \dfrac{n}{n+1}\theta = \theta - \dfrac{\theta}{n+1}$, i.e., the estimate $\hat{\theta}_n$ is asymptotically unbiased with $\mathbf{D}_\theta \hat{\theta}_n = \dfrac{n}{(n+1)^2(n+2)} \theta^2 \to 0$ as $n \to \infty$ and, consequently, $\hat{\theta}_n$ is consistent. The distribution function for $X_{(n)}$ was found in Problem 2.79. For $t \geqslant 0$ we have

$$\mathbf{P}_\theta\left(\frac{\theta - \hat{\theta}_n}{\theta} n \leqslant t\right) = \mathbf{P}_\theta\left(X_{(n)} \geqslant \theta\left(1 - \frac{t}{n}\right)\right) = 1 - \left(1 - \frac{t}{n}\right)^n \to 1 - e^{-t};$$

i.e., the distribution of $\hat{\theta}_n$ is not asymptotically normal (the model is not regular).

2.101. The form of the likelihood function $L(\mathbf{x}; \theta) = I(\theta + 1/2 \geqslant x_{(n)}) \times I(x_{(1)} \geqslant \theta - 1/2)$ implies that $L(\mathbf{x}; \theta) = 1 \; \forall \theta \in [x_{(n)} - 1/2, x_{(1)} + 1/2]$. Consequently, any θ from this interval maximizes $L(\mathbf{x}; \theta)$. An arbitrary point

$T \in [X_{(n)} - 1/2, X_{(1)} + 1/2]$ can be written in the form $T = \alpha(X_{(n)} - 1/2) + (1 - \alpha)(X_{(1)} + 1/2)$, $\alpha \in [0, 1]$. From this we find the condition

$$\mathbf{E}_\theta T = \frac{1}{2} - \alpha + \alpha \mathbf{E}_\theta X_{(n)} + (1 - \alpha)\mathbf{E}_\theta X_{(1)} = \theta \ \forall\theta$$

from which we obtain an unbiased estimator for θ. Using the solution to Problem 1.36, we get $\alpha = 1/2$, i.e., $T = (1/2)(X_{(1)} + X_{(n)})$ is the midpoint of the interval.

2.102. Here the density is $f(x; \theta) = b^{-\alpha}\alpha(x - \theta)^{\alpha-1} \exp[-b^{-\alpha}(x - \theta)^\alpha]$, $x \geqslant \theta$, and the likelihood function has the form

$$L(\mathbf{x}; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

$$= b^{-\alpha n}\alpha^n I(x_{(1)} \geqslant \theta) \prod_{i=1}^{n} (x_i - \theta)^{\alpha-1} \exp\left\{-\frac{(x_i - \theta)^\alpha}{b^\alpha}\right\}.$$

It is monotone increasing on the interval $-\infty < \theta \leqslant x_{(1)}$ and is zero for $\theta > x_{(1)}$. Consequently, it has a maximum at the point $\theta = x_{(1)}$. The model under investigation is not regular, but its m.l.e. $\hat{\theta}_n = X_{(1)}$ is, by Problem 2.26, asymptotically unbiased and consistent. The asymptotic distribution of $X_{(1)}$ is given in the solution to Problem 1.37 and it is not normal.

2.103. Here the likelihood function is $L(\mathbf{x}; \theta) = \left(\frac{2}{\theta}\right)^n e^{-T/\theta} \prod_{i=1}^{n} x_i$, $T = \sum_{i=1}^{n} x_i^2$, and the likelihood equation $\frac{\partial \ln L}{\partial \theta} = 0$ has the only solution $\hat{\theta}_n = T/n$ which maximizes $L(\mathbf{x}; \theta)$. The model is a special case of Weibull's distribution with an unknown scale parameter (see Problem 2.76). Therefore, $\hat{\theta}_n$ coincides with the complete sufficient statistic and is an unbiased optimum estimate for θ.

2.104. The form of $\hat{\tau}_n$ follows from the solution to Problem 2.84 and the invariance of maximum likelihood estimates. From Problem 2.21 we get $\hat{\tau}_n = \frac{\lambda n}{\lambda n - 1} \tau_1^*$, where $\tau_1^* = \frac{\lambda n - 1}{T}$ is an unbiased estimator for θ^{-1}. We thus have

$$\mathbf{E}_\theta \hat{\tau}_n = \frac{\lambda n}{\lambda n - 1}\theta^{-1} = \theta^{-1} + \frac{1}{(\lambda n - 1)\theta},$$

i.e., $\hat{\tau}_n$ is an asymptotically unbiased estimate.

Further,

$$\mathbf{D}_\theta \hat{\tau}_n = \left(\frac{\lambda n}{\lambda n - 1}\right)^2 \mathbf{D}_\theta \tau_1^* = \left(\frac{\lambda n}{\lambda n - 1}\right)^2 [(\lambda n - 1)^2 \mathbf{E}_\theta T^{-2} - \theta^{-2}]$$
$$= \frac{(\lambda n)^2}{(\lambda n - 1)^2(\lambda n - 2)} \theta^{-2} \sim \frac{\theta^{-2}}{\lambda n} = \frac{(\tau'(\theta))^2}{ni(\theta)}$$

(see Problem 2.43). We see that $\hat{\tau}_n$ is consistent and

$$\mathcal{L}_\theta(\hat{\tau}_n) \sim \mathcal{N}\left(\frac{1}{\theta}, \frac{1}{\lambda n \theta^2}\right).$$

2.105. In order to maximize the likelihood function, we must minimize the sum $\sum_{i=1}^{n} |x_i - \theta| = \sum_{i=1}^{n} |x_{(i)} - \theta|$, whence follows that $\hat{\theta}_n$ coincides with the sample median. We cannot use here the theorems on the asymptotic normality of m.l.e.'s or sample quantiles because the density $f(x; \theta)$ is not differentiable at the point θ. Nevertheless, we have $\mathcal{L}_\theta(\hat{\theta}_n) \sim \mathcal{N}(\theta, 1/n)$ as was the case in Problem 1.32.

2.106. We seek the limiting distribution for the estimator T_n as $n \to \infty$. We have

$$\mathbf{P}_\theta(\sqrt{n}(T_n - \theta) \leqslant x) = p_n \mathbf{P}_\theta(\sqrt{n}(\overline{X} - \theta) \leqslant x) + (1 - p_n)\mathbf{P}_\theta(\sqrt{n}(b\overline{X} - \theta) \leqslant x),$$

where

$$p_n = \mathbf{P}_\theta(|\overline{X}| \geqslant a_n) = \Phi(\sqrt{n}(\theta - a_n)) + \Phi(-\sqrt{n}(\theta + a_n)) \to \begin{cases} 1 & \text{for} \quad |\theta| > 0, \\ 0 & \text{for} \quad \theta = 0 \end{cases}$$

as $n \to \infty$. Thus,

$$\mathcal{L}_\theta(\sqrt{n}(T_n - \theta)) \to \begin{cases} \mathcal{N}(0, 1) & \text{for} \quad |\theta| > 0, \\ \mathcal{N}(0, b^2) & \text{for} \quad \theta = 0 \end{cases}$$

as $n \to \infty$. From this we find

$$\operatorname{eff}(T_n; \theta) = \begin{cases} 1 & \text{for} \quad |\theta| > 0, \\ b^{-2} & \text{for} \quad \theta = 0, \end{cases}$$

i.e., $\operatorname{eff}(T_n; \theta) \geqslant 1$ for $|b| < 1$, the strict inequality holding at the point $\theta = 0$, which is, consequently, the point of superefficiency.

2.107. We know that in the model $R(0, \theta)$ (see the solution to Problem 2.100) $\hat{\theta}_n = X_{(n)}$ and $\mathbf{D}_\theta \hat{\theta}_n = \theta(n^{-2})$. In Weibull's model (see the solutions to Problems 2.102 and 1.37) $\hat{\theta}_n = X_{(1)}$ and $\mathbf{D}_\theta \hat{\theta}_n = \theta(n^{-2/\alpha})$ for $0 < \alpha \leqslant 1$, i.e., the variance of the m.l.e.'s may be as small as possible depending on the value of α.

2.108. Here (see the solution to Problem 2.84) $\hat{\theta}_n = \overline{X}$ and by virtue of the invariance we have $\hat{\tau}_n = \overline{X}^{-1}$. But we know (see Problem 1.39) that $\mathcal{L}_\theta(n\overline{X}) = \Pi(n\theta)$ and $\mathbf{P}_\theta(\overline{X} = 0) = e^{-n\theta} > 0$. Consequently, the random varia-

ble $\overline{X}$ becomes zero with a positive probability for any n, and the statistic $\hat{\tau}_n$ has no finite moments. At the same time its asymptotic variance is $[\tau'(\theta)]^2/[ni(\theta)] = (\theta^3 n)^{-1}$ (see Problem 2.43).

2.109. The asymptotic variance of the m.l.e. $\hat{\tau}_n$ is independent of the parameter θ if and only if the function $\tau(\theta)$ satisfies the relation $\sigma_\tau^2(\theta) = (\tau'(\theta))^2/i(\theta) \equiv \text{const}$. From this and Problem 2.43 we find the respective equation $\tau'(\theta) = c/\sqrt{\theta(1-\theta)}$ for the model $Bi(k, \theta)$. Its solution is $\tau(\theta) = \arcsin\sqrt{\theta}$ (up to a constant factor) for which $\sigma_\tau^2(\theta) = 1/(4k)$. For the model $\Pi(\theta)$ the sought-for function is the solution of the equation $\tau'(\theta) = c/\sqrt{\theta}$, i.e., $\tau(\theta) = \sqrt{\theta}$ and $\sigma_\tau^2(\theta) = 1/4$. For the models $\mathcal{N}(\mu, \theta^2)$ and $\Gamma(\theta, \lambda)$ we have the equation $\tau'(\theta) = c/\theta$, i.e., $\tau(\theta) = \ln\theta$. Thus, using Problem 2.84, we have the following useful approximations:

for the model $Bi(k, \theta)$: $\mathcal{L}_\theta(\arcsin\sqrt{\hat{\theta}_n}) \sim \mathcal{N}\left(\arcsin\sqrt{\theta}, \dfrac{1}{4kn}\right)$, $\hat{\theta}_n = \overline{X}/k$;

for the model $\Pi(\theta)$: $\mathcal{L}_\theta(\sqrt{\hat{\theta}_n}) \sim \mathcal{N}\left(\sqrt{\theta}, \dfrac{1}{4n}\right)$, $\hat{\theta}_n = \overline{X}$;

for the model $\mathcal{N}(\mu, \theta^2)$: $\mathcal{L}_\theta(\ln\hat{\theta}_n) \sim \mathcal{N}\left(\ln\theta, \dfrac{1}{2n}\right)$, $\hat{\theta}_n = \left[\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2\right]^{1/2}$;

for the model $\Gamma(\theta, \lambda)$: $\mathcal{L}_\theta(\ln\hat{\theta}_n) \sim \mathcal{N}\left(\ln\theta, \dfrac{1}{\lambda n}\right)$, $\hat{\theta}_n = \overline{X}/\lambda$.

2.111. From the solution to Problem 2.83 we find that for the given $\eta = k$ (the number of elements observed in the sample) by maximizing the likelihood function in $N \geqslant k$, we maximize the functions $g(k; N) = C_N^k N^{-n}$ in N or, which is equivalent, the functions

$$f(N) = \sum_{j=0}^{k-1} \ln(N - j) - n\ln N, \quad N \geqslant k.$$

Consider the difference

$$\Delta f(N) = f(N+1) - f(N) = \ln\frac{N+1}{N-k+1} - n\ln\frac{N+1}{N}$$
$$\equiv [S(N, k) - n]\ln\frac{N+1}{N}.$$

Here the function $S(N, k)$ is monotone decreasing in N for $k > 1$. Indeed, if we introduce the function $\varphi(x) = \ln\left(1 - \dfrac{x}{N+2}\right)\Big/\ln\left(1 - \dfrac{x}{N+1}\right)$, $0 < x < N + 1$, the inequality $S(N, k) > S(N+1, k)$ will be equivalent to $\varphi(k) < \varphi(1)$. The function $\varphi(x)$ is monotone decreasing because the inequality $\varphi'(x) < 0$ reduces to

$$(N + 1 - x)\ln\left(1 - \frac{x}{N+1}\right) > (N + 2 - x)\ln\left(1 - \frac{x}{N+2}\right).$$

This follows from the fact that the function $\psi(y) = (y - x)\ln\left(1 - \frac{x}{y}\right)$, $y > x$, is monotone decreasing $\left(\psi'(y) = \ln\left(1 - \frac{x}{y}\right) + \frac{x}{y} < 0\right)$. Thus, for the given values of n and $k > 1$ the inequalities $S(N, k) \leqslant n < S(N - 1, k)$ uniquely define the integer $N_0 = N_0(k, n)$. Here $\Delta f(N) > 0$ for $N \leqslant N_0 - 1$ and $\Delta f(N_0) \leqslant 0$, $\Delta f(N) < 0$ for $N \geqslant N_0 + 1$. This means that the function $f(N)$ is monotone increasing for $N \leqslant N_0$ and monotone decreasing for $N \geqslant N_0$ if $S(N_0, k) \neq n$. If $S(N_0, k) = n$, then $f(N_0) = f((N_0 + 1)$. To the left of N_0 and to the right of $N_0 + 1$ we have $f(N) < f(N_0)$. Thus, in any case $\max_N f(N) = f(N_0)$, q.e.d. Now let $k = 1$. Then $g(1; N) = 1/N^{n-1}$ and maximum is attained at $N = 1$.

We have $\hat{N} = \eta$ if and only if there holds the condition $S(\eta, \eta) \leqslant n$ (since $S(\eta - 1, \eta) = \infty$ by the definition of the function S). This condition can be written as $\ln(\eta + 1) \leqslant n \ln \frac{\eta + 1}{\eta}$.

In the given asymptotic conditions we find an approximate solution for $\hat{\alpha}$ from the relation

$$-\ln\left(1 - \frac{\eta}{n}\hat{\alpha}\right) = n \ln \frac{\hat{N} + 1}{\hat{N}} \simeq \hat{\alpha}.$$

Then $\frac{\eta}{n} = a(\hat{\alpha})$, where $a(\alpha) = \frac{1 - e^{-\alpha}}{\alpha}$. We use $a^{-1}(t)$ to denote the function inverse to $a(\alpha)$ and obtain $\hat{\alpha} = a^{-1}(\eta/n)$.

For an arbitrary value of m we get the function $g_m(k; N) = C_N^k (C_N^m)^{-n}$ instead of the function $g(k; N)$. We maximize the former in N and find that for $\eta > m$ the m.l.e. $\hat{N}_m$ is defined by the inequalities

$$S_m(\hat{N}_m, \eta) \leqslant n < S_m(\hat{N}_m - 1, \eta),$$

where

$$S_m(N, k) = \ln \frac{N + 1}{N + 1 - k} \Big/ \ln \frac{N + 1}{N + 1 - m} \text{ for } N \geqslant k > m,$$

$$S_m(k - 1, k) = \infty.$$

If $\eta = m$, we have $\hat{N}_m = m$.

2.112. If $\mu_2 = k$, then $\hat{N}$ is the value of N which maximizes the likelihood function

$$\mathbf{P}_N(\mu_2 = k) = C_{m_1}^k C_{N-m_1}^{m_2-k} \Big/ C_N^{m_2} \equiv g(N).$$

Here $\frac{g(N)}{g(N-1)} = \frac{(N - m_1)(N - m_2)}{N(N - m_1 - m_2 + k)}$, whence follows that the inequalities $g(N) \gtreqless g(N - 1)$ are equivalent to $Nk \lesseqgtr m_1 m_2$. We write $N_0 = [m_1 m_2/k]$ and find that for $N \leqslant N_0$ the function $g(N)$ is increasing, while for $N \geqslant N_0$ it is decreasing if the number $m_1 m_2/k$ is not integer, i.e., in this case N_0 is

the point of maximum. If $N_0 = m_1m_2/k$, maximum is attained at two points, i.e., $N_0 - 1$ and N_0. Thus, the value of $\hat{N}$ coincides with N_0 in any case. We established in Problem 2.36 that (for $n = 2$, $m_1 = m_2 = m$) the only unbiased estimator for $\tau(N) = 1/N$, which is a linear function of μ_2, has the form μ_2/m^2. The maximum likelihood estimate is $\hat{\tau} = 1/\hat{N} = 1\Big/\left[\dfrac{m^2}{\mu_2}\right] = \dfrac{\mu_2}{m^2}\left(1 - \dfrac{\varepsilon\mu_2}{m^2}\right)^{-1}$,

where $\varepsilon = \dfrac{m^2}{\mu_2} - \left[\dfrac{m^2}{\mu_2}\right]$. Clearly, it is a biased estimate.

2.113. (1) We first prove that d_n is a sufficient statistic with the help of the factorization test, i.e., we show that the representation $\mathbf{P}_D(\mathbf{X} = \mathbf{x}) = g(d_n; D)h(\mathbf{x})$, $d_n = \sum_{i=1}^{n} x_i$, is valid for the likelihood function $\mathbf{P}_D(\mathbf{X} = \mathbf{x})$, $\mathbf{x} = (x_1, \ldots, x_n)$, $x_i = 0, 1$, $i = 1, \ldots, n$. Using the formula for the product of probabilities, we may write

$$\mathbf{P}_D(\mathbf{X} = \mathbf{x}) = \mathbf{P}_D(X_1 = x_1)\mathbf{P}_D(X_2 = x_2|X_1 = x_1)$$
$$\times \ldots \mathbf{P}_D(X_n = x_n|X_1 = x_1, X_2 = x_2, \ldots, X_{n-1} = x_{n-1}).$$

We then have

$$\mathbf{P}_D(X_j = x_j|X_1 = x_1, \ldots, X_{j-1} = x_{j-1})$$
$$= \begin{cases} \dfrac{D - x_1 - \ldots - x_{j-1}}{N - j + 1} & \text{for } x_j = 1, \\[2ex] 1 - \dfrac{D - x_1 - \ldots - x_{j-1}}{N - j + 1} & \text{for } x_j = 0 \end{cases}$$
$$= (D - x_1 - \ldots - x_{j-1})^{x_j}(N - D - j + 1 + x_1 + \ldots + x_{j-1})^{1-x_j}/(N - j + 1)$$

and

$$\mathbf{P}_D(\mathbf{X} = \mathbf{x}) = D^{x_1}(D - x_1)^{x_2}\ldots(D - x_1 - \ldots - x_{n-1})^{x_n}(N - D)^{1-x_1}$$
$$\times (N - D - (1 - x_1))^{1-x_2}$$
$$\times \ldots (N - D - (n - 1 - x_1 - \ldots - x_{n-1}))^{1-x_n}/(N)_n.$$

Using induction on n, we write the formula

$$M^{\varepsilon_1}(M - \varepsilon_1)^{\varepsilon_2}\ldots(M - \varepsilon_1 - \ldots - \varepsilon_{n-1})^{\varepsilon_n} = M(M-1)\ldots\left(M - \sum_{j=1}^{n}\varepsilon_j + 1\right)$$
$$= M!\Big/\left(M - \sum_{j=1}^{n}\varepsilon_j\right)!,$$

where $\varepsilon_j = 0, 1$, which finally gives us

$$\mathbf{P}_D(\mathbf{X} = \mathbf{x}) = \frac{D!(N - D)!(N - n)!}{(D - d_n)!(N - D - n + d_n)!N!} = C_{N-n}^{D-d_n}/C_N^D.$$

We show that d_n is a complete statistic. We must then prove that if $\mathbf{E}_D\varphi(d_n) = 0$ for $D = 0, 1, \ldots, N$, then $\varphi(k) = 0$, $k \in [0, 1, \ldots, \min(D, n)]$ for all possible D, i.e., for $k = 0, 1, \ldots, n$. Since $\mathscr{L}_D(d_n) = H(D, N, n)$, the previous condition has the form

$$\sum_{k=0}^{\min(D,n)} \varphi(k)C_D^k C_{N-D}^{n-k}/C_N^n = 0,\ D = 0, 1, \ldots, N.$$

We put $D = 0$, $D = 1$, etc. to obtain $\varphi(0) = 0$, $\varphi(1) = 0$, etc.

Now let $\tau(D)$ be the given estimated function of the parameter D. Its optimum estimator is the solution of the unbiasedness equation

$$\mathbf{E}_D T(d_n) = \tau(D),\ D = 0, 1, \ldots, N. \qquad (*)$$

For any function T the left-hand side is a polynomial in D of degree $\leqslant n$. If $\tau(D)$ is not a polynomial of degree $\leqslant n$, then the equation $(*)$ has no solution, and there are no unbiased estimators for such functions.

Finally, we assume that $\tau(D)$ is the polynomial from the statement of the problem. We check whether τ^* meets the condition $(*)$. We have

$$\mathbf{E}_D\tau^* = \sum_{k=0}^{n} T(k)f(k;\ D, n) = \sum_{j=0}^{N} a_j \frac{(N)_j}{(n)_j} \sum_{k=j}^{n} (k)_j f(k;\ D, n).$$

Here

$$\sum_{k=j}^{n} (k)_j f(k;\ D, n) = \mathbf{E}_D(d_n)_j = (D)_j(n)_j/(N)_j.$$

Therefore,

$$\mathbf{E}_D\tau^* = \sum_{j=0}^{n} a_j(D)_j = \tau(D)\ \forall D.$$

Consequently, τ^* is an optimum estimator for $\tau(D)$ as a function of a complete sufficient statistic.

(2) The function $\tau_1(D)$ is a polynomial of degree one. Therefore, its unbiased estimator always exists and has the form $\tau_1^* = Nd_n/n$.

The function $\tau_2(D) = (N - 1)D - (D)_2$ is a polynomial of degree two. Therefore, it only has an unbiased estimator for a sample of size $n \geqslant 2$. From the above we find that $\tau_2^* = d_n(n - d_n)(N)_2/(n)_2$.

(3) In order to find a m.l.e. for the parameter D, we must maximize the quantity $g(d_n;\ D) = C_{N-n}^{D-d_n}/C_N^D$. But

$$\frac{g(d_n;\ D + 1)}{g(d_n;\ D)} = \frac{(D + 1)(N - n + d_n - D)}{(D + 1 - d_n)(N - D)},$$

whence we obtain $g(d_n;\ D + 1) \gtreqless g(d_n;\ D)$ for $D \lesseqgtr \dfrac{d_n}{n}(N + 1) - 1$. If the

number $\frac{d_n}{n}(N+1)$ is not integer, the maximum is attained at the point $D_0 = \left[\frac{d_n}{n}(N+1)\right]$. Otherwise, maximum is attained at two points, i.e., D_0 and $D_0 - 1$. Thus, $\hat{D}_n = D_0$ in any case.

2.114. The likelihood function of the jth sample has the form

$$L_j = L_j(\mathbf{x}_j;\ \theta_{j_1}, \theta_2) = \frac{1}{(\sqrt{2\pi}\,\theta_2)^{n_j}} \exp\left\{-\frac{n_j}{2\theta_2^2}(s_j^2 + (\bar{x}_j - \theta_{j_1})^2)\right\}$$

(see the solution to Problem 2.86). Since the samples are independent, the likelihood function is

$$L = \prod_{j=1}^{k} L_j = (2\pi e t^2)^{-n/2} \exp\left\{-\frac{1}{2\theta_2^2}\sum_{j=1}^{k} n_j(\bar{x}_j - \theta_{j_1})^2 - n\left(\frac{1}{2}\left(\frac{t^2}{\theta_2^2} - 1\right) - \ln\frac{t}{\theta_2}\right)\right\}$$

for all of them. Here $t^2 = \frac{1}{n}\sum_{j=1}^{k} n_j s_j^2$, $n = n_1 + \ldots + n_k$. The power of the exponent is non-positive and becomes zero only for $\theta_{j_1} = \bar{x}_j$, $j = 1, \ldots, k$, $\theta_2 = t$. Thus the m.l.e.'s of the parameters have the indicated form.

We now have (see the solution to Problem 2.1)

$$\mathbf{E}_\theta \hat{\theta}_2^2 = \frac{1}{n}\sum_{j=1}^{k} n_j \mathbf{E}_\theta S_j^2 = \frac{1}{n}\sum_{j=1}^{k} n_j \frac{n_j - 1}{n_j}\theta_2^2 = \frac{n-k}{n}\theta_2^2.$$

Consequently, the statistic $\frac{n}{n-k}\hat{\theta}_2^2$ is an unbiased estimator for θ_2^2.

2.115. Taking into account the choice of c_γ, we have

$$\gamma = \mathbf{P}_\theta(\sqrt{n}|\overline{X} - \theta|/\theta < c_\gamma) = \mathbf{P}_\theta(\theta(1 - c_\gamma/\sqrt{n}) < \overline{X} < \theta(1 + c_\gamma/\sqrt{n}))$$
$$= \mathbf{P}_\theta(\overline{X}/(1 + c_\gamma/\sqrt{n}) < \theta < \overline{X}/(1 - c_\gamma/\sqrt{n})).$$

For the model with a negative parameter the original equation becomes $\gamma = \mathbf{P}_\theta(\sqrt{n}|\overline{X} - \theta|/|\theta| < c_\gamma)$, whence the sought-for interval is $(\overline{X}/(1 - c_\gamma/\sqrt{n}), \overline{X}/(1 + c_\gamma/\sqrt{n}))$.

2.116. (1) Here $\mathscr{L}_\theta(G(\mathbf{X};\theta)) = \mathscr{N}(0, 1)$ and $G(\mathbf{X};\theta)$ is a central statistic. We get

$$\mathbf{P}_\theta(g_1 < G(\mathbf{X};\theta) < g_2) = \Phi(g_2) - \Phi(g_1) = \gamma,$$

and $T_1 = \overline{X} - \frac{\sigma}{\sqrt{n}}g_2$, $T_2 = \overline{X} - \frac{\sigma}{\sqrt{n}}g_1$ are the solutions to $G(\mathbf{X};\theta) = g_1, g_2$.

Thus, the confidence interval $\Delta_\gamma(\mathbf{X})$ has the indicated form. Its length is $l = \frac{\sigma}{\sqrt{n}}(g_2 - g_1)$, and therefore in order to construct the shortest interval, we should minimize the difference $g_2 - g_1$ under the condition $\Phi(g_2) - \Phi(g_1) = \gamma$. We use Lagrange's method for finding the conditional extremum. We form the Lagrange function

$$H(g_1, g_2, \lambda) = g_2 - g_1 + \lambda(\Phi(g_2) - \Phi(g_1) - \gamma)$$

and equate all its partial derivatives to zero to obtain the system of equations $\Phi'(g_1) = \Phi'(g_2)$, $\Phi(g_2) - \Phi(g_1) = \gamma$. Since the function $\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is even, the first equation gives $g_1 = -g_2$. Taking into account the second equation and the relation $\Phi(-x) = 1 - \Phi(x)$, we obtain $2\Phi(g_2) - 1 = \gamma$, whence $g_2 = c_\gamma$.

(2) The length l of the confidence interval $\Delta_\gamma^*(\mathbf{X})$, the confidence interval γ, and the sample size n are related here as $l = \frac{2\sigma c_\gamma}{\sqrt{n}}$. Then for the given l and γ the required number of observations is $n = n(l, \gamma) = \left[4\frac{\sigma^2 c_\gamma^2}{l^2}\right]$, and for the given n and l the confidence level is $\gamma = \gamma(n, l) = 2\Phi\left(\frac{l\sqrt{n}}{2\sigma}\right) - 1$. Specifically, $c_{0.99} = 2.5758$ and $n = 106$ for $l = 0.5$ (at $\sigma = 1$), while $n = 2653$ for $l = 0.1$.

2.117. The distribution density of the random variable T/θ, which in our case is a central statistic, is

$$\frac{\partial}{\partial x}\mathbf{P}_\theta\left(\frac{T}{\theta} \leqslant x\right) = \frac{\partial}{\partial x}\mathbf{P}_\theta\left(\frac{T^2}{\theta^2} \leqslant x^2\right) = 2xk_n(x^2).$$

Therefore,

$$\mathbf{P}_\theta(\theta \in \delta_\gamma(\mathbf{X})) = \mathbf{P}_\theta\left(a_1 < \frac{T}{\theta} < a_2\right) = 2\int_{a_1}^{a_2} xk_n(x^2)\,dx = \gamma.$$

The shortest of the intervals under consideration is found by minimizing the ratio a_2/a_1 under the condition $\int_{a_1}^{a_2} xk_n(x^2)\,dx = \gamma/2$. The method of Lagrange multipliers gives

$$a_1^2 k_n(a_1^2) = a_2^2 k_n(a_2^2), \quad \int_{a_1^2}^{a_2^2} k_n(x)\,dx = \gamma.$$

We write $a_1^2 = \chi^2_{\alpha_1,n}$, $a_2^2 = \chi^2_{1-\alpha_2,n}$ and reduce the above equations to

$$\chi^2_{1-\alpha_2,n}/\chi^2_{\alpha_1,n} = \exp\left\{\frac{1}{n}(\chi^2_{1-\alpha_2,n} - \chi^2_{\alpha_1,n})\right\}, \quad \alpha_1 + \alpha_2 = 1 - \gamma.$$

These relations uniquely define α_1^* and α_2^* and thus $\chi^2_{\alpha_1^*,n}$, $\chi^2_{1-\alpha_2^*,n}$ [7, Table 2.3]. Thus, the optimal interval has the form

$$\delta_\gamma^*(\mathbf{X}) = (T_1^*, T_2^*) = (T/\chi_{1-\alpha_2^*,n}, T/\chi_{\alpha_1^*,n}).$$

2.118. Here the central statistic is $G(\mathbf{X}; \tau) = T^2/\tau$, $\tau = \theta^2$, and the assertion follows from the previous problem.

2.119. From the hint we have $\mathbf{P}_\theta(\sqrt{n-1}(\overline{X} - \theta_1)/S \leqslant t_{\gamma,n-1}) = \gamma$, whence we find the lower γ-confidence interval $(\overline{X} - t_{\gamma,n-1}S/\sqrt{n-1} < \theta_1)$. Similarly, the relation $\mathbf{P}_\theta(\sqrt{n-1}(\overline{X} - \theta_1)/S \geqslant -t_{\gamma,n-1}) = \gamma$ gives the upper γ-confidence interval $(\theta_1 < \overline{X} + t_{\gamma,n-1}S/\sqrt{n-1})$. The central two-sided γ-confidence interval is $(\overline{X} \pm t_{(1+\gamma)/2,n-1}S/\sqrt{n-1})$. This is the shortest interval among all the γ-confidence intervals of the form $(\overline{X} - a_1S, \overline{X} + a_2S)$, which is proved similarly to the assertion from Problem 2.116.

2.120. Here nS^2/τ is a central statistic, and

$$(nS^2/\chi^2_{(1+\gamma)/2,n-1},\ nS^2/\chi^2_{(1-\gamma)/2,n-1})$$

is the central γ-confidence interval. The lower and upper γ-confidence intervals are $(nS^2/\chi^2_{\gamma,n-1} < \theta_2^2)$ and $(\theta_2^2 < nS^2/\chi^2_{1-\gamma,n-1})$, respectively.

2.122. The sample means $\overline{X}$ and $\overline{Y}$ are independent and normal as $\mathcal{N}(\theta^{(1)}, \sigma_1^2/n)$ and $\mathcal{N}(\theta^{(2)}, \sigma_2^2/m)$, respectively. Therefore, $\mathcal{L}(\overline{X} - \overline{Y}) = \mathcal{N}(\tau, \sigma^2)$, i.e., $(\overline{X} - \overline{Y} - \tau)/\sigma$ is a central statistic for τ. The γ-confidence interval is sought as in Problem 2.116 and has the form $(\overline{X} - \overline{Y} \pm c_\gamma\sqrt{\sigma_1^2/n + \sigma_2^2/m})$.

2.123. We have $\mathcal{L}\left(\dfrac{\overline{X} - \theta_1^{(1)}}{\theta_2}\right) = \mathcal{N}\left(0, \dfrac{1}{n}\right)$, $\mathcal{L}\left(\dfrac{\overline{Y} - \theta_1^{(2)}}{\theta_2}\right) = \mathcal{N}\left(0, \dfrac{1}{m}\right)$, $\mathcal{L}\left(\dfrac{n}{\theta_2^2}S^2(\mathbf{X})\right) = \chi^2(n-1)$, $\mathcal{L}\left(\dfrac{m}{\theta_2^2}S^2(\mathbf{Y})\right) = \chi^2(m-1)$. Since the samples are independent, we also have

$$\mathcal{L}\left(\frac{\overline{X} - \overline{Y} - \tau}{\theta_2}\right) = \mathcal{N}\left(0, \frac{1}{n} + \frac{1}{m}\right) \text{ or } \mathcal{L}\left(\sqrt{\frac{mn}{m+n}}\,\frac{\overline{X} - \overline{Y} - \tau}{\theta_2}\right) = \mathcal{N}(0, 1),$$

$$\mathcal{L}\left(\frac{1}{\theta_2^2}[nS^2(\mathbf{X}) + mS^2(\mathbf{Y})]\right) = \chi^2(m + n - 2).$$

Here the random variables $\overline{X} - \overline{Y}$ and $nS^2(\mathbf{X}) + mS^2(\mathbf{Y})$ are independent (by Fisher's theorem). Then $\mathcal{L}(t_{m+n-2}) = S(m + n - 2)$, and, consequently, t_{m+n-2} is a central statistic for τ. The respective γ-confidence interval is constructed as in Problem 2.119 and has the form

$$\left(\overline{X} - \overline{Y} \pm t_{(1+\gamma)/2,m+n-2}\left[\frac{m+n}{mn(m+n-2)}(nS^2(\mathbf{X}) + mS^2(\mathbf{Y}))\right]^{1/2}\right).$$

2.125. By Fisher's theorem we have $\mathscr{L}(nS^2(\mathbf{X})/\theta_2^{(1)^2}) = \chi^2(n-1)$, $\mathscr{L}(mS^2(\mathbf{Y})/\theta_2^{(2)^2}) = \chi^2(m-1)$, and $S^2(\mathbf{X})$ and $S^2(\mathbf{Y})$ are independent. Consequently, $\mathscr{L}(F_{n-1,m-1}) = S(n-1, m-1)$. Then the central γ-confidence interval for τ has the form

$$\left(\frac{n(m-1)}{m(n-1)}\frac{S^2(\mathbf{X})}{S^2(\mathbf{Y})}\Big/F_{(1+\gamma)/2,n-1,m-1},\ \frac{n(m-1)}{m(n-1)}\frac{S^2(\mathbf{X})}{S^2(\mathbf{Y})}\Big/F_{(1-\gamma)/2,n-1,m-1}\right).$$

1.127. Since $\mathscr{L}(2n\overline{X}/\theta_1) = \chi^2(2n)$, $\mathscr{L}(2m\overline{Y}/\theta_2) = \chi^2(2m)$, we have $\mathscr{L}(\tau\overline{X}/\overline{Y}) = S(2n, 2m)$. As in the previous problem, we find the sought-for interval

$$(F_{(1-\gamma)/2,2n,2m}\overline{Y}/\overline{X},\ F_{(1+\gamma)/2,2n,2m}\overline{Y}/\overline{X}).$$

2.128. Since $\mathbf{P}_\theta(X_{(1)} > x) = \mathbf{P}_\theta^n(X_1 > x) = e^{-n(x-\theta)}$ for $x \geqslant \theta$, we have $\mathbf{P}_\theta(\theta \leqslant X_{(1)} \leqslant x + \theta) = 1 - e^{-nx} = \gamma$ for $x = -\frac{1}{n}\ln(1-\gamma)$, i.e., the sought-for interval is

$$\left(X_{(1)} + \frac{1}{n}\ln(1-\gamma),\ X_{(1)}\right).$$

2.129. Since $\mathscr{L}_\theta(X_i/\theta) = R(0, 1)$, we have $\mathscr{L}_\theta(X_{(n)}/\theta) = B(n, 1)$ (see Problem 1.35). Then for $0 \leqslant t \leqslant 1$ we have

$$\mathbf{P}_\theta((X_{(n)}/\theta)^n \leqslant t) = \mathbf{P}_\theta(X_{(n)}/\theta \leqslant t^{1/n}) = n\int_0^{t^{1/n}} x^{n-1}\,dx = t.$$

Whence

$$\mathbf{P}_\theta(X_{(n)} \leqslant \theta \leqslant X_{(n)}/\sqrt[n]{1-\gamma}) = \mathbf{P}_\theta((X_{(n)}/\theta)^n \geqslant 1-\gamma) = \gamma.$$

2.130. Since $\mathscr{L}_\theta(2T/\theta^\lambda) = \chi^2(2n)$, we have

$$\mathbf{P}_\theta\left(\chi^2_{(1-\gamma)/2,2n} < \frac{2T}{\tau(\theta)} < \chi^2_{(1+\gamma)/2,2n}\right) = \gamma,$$

which is equivalent to our assertion.

2.131. We have

$$\begin{aligned}
&\mathbf{P}_\theta((\theta_1, \tau) \in \mathscr{S}_\gamma(\mathbf{X}))\\
&= \mathbf{P}_\theta(\sqrt{n}|\overline{X} - \theta_1|\theta_2^{-1} < c_{\gamma_1},\ \chi^2_{(1-\gamma_2)/2,n-1} < nS^2/\theta_2^2 < \chi^2_{(1+\gamma_2)/2,n-1})\\
&= \mathbf{P}_\theta(\sqrt{n}|\overline{X} - \theta_1|\theta_2^{-1} < c_{\gamma_1})\mathbf{P}_\theta(\chi^2_{(1-\gamma_2)/2,n-1} < nS^2/\theta_2^2 < \chi^2_{(1+\gamma_2)/2,n-1})\\
&= (\Phi(c_{\gamma_1}) - \Phi(-c_{\gamma_1}))\left(\frac{1+\gamma_2}{2} - \frac{1-\gamma_2}{2}\right) = \gamma_1\gamma_2 = \gamma.
\end{aligned}$$

2.132. According to Problem 1.59 (b), the quadratic form

$$Q = \frac{n}{1-\varrho^2}\left[\frac{1}{\sigma_1^2}(\overline{X}_1 - \theta_1)^2 - \frac{2\varrho}{\sigma_1\sigma_2}(\overline{X}_1 - \theta_1)(\overline{X}_2 - \theta_2) + \frac{1}{\sigma_2^2}(\overline{X}_2 - \theta_2)^2\right]$$

is distributed as $\chi^2(2)$ for any θ and Σ. Therefore,

$$\gamma = \mathbf{P}_\theta(Q \leqslant \chi^2_{\gamma,2}) = \mathbf{P}_\theta(\theta \in \mathscr{I}_\gamma(\mathbf{X})),$$

where

$$\mathscr{I}_\gamma(\mathbf{X}) = \{\theta\colon Q = n(\overline{X_1} - \theta_1, \overline{X_2} - \theta_2)'\Sigma^{-1}(\overline{X_1} - \theta_1, \overline{X_2} - \theta_2) \leqslant \chi^2_{\gamma,2}\}.$$

Thus, $\mathscr{I}_\gamma(\mathbf{X})$ is the sought-for γ-confidence region for $\theta = (\theta_1, \theta_2)$. This is the inside of an ellipse with centre at a random point $(\overline{X_1}, \overline{X_2})$ whose boundary is defined by the equation $Q = \chi^2_{\gamma,2}$.

2.133. Since $\mathscr{L}_\theta(nT) = Bi(n, \theta)$ (see Problem 1.39 (3)), the random variable T assumes the values of $0, 1/n, 2/n, \ldots, n/n$ and its distribution function

$$F\left(\frac{k}{n}; \theta\right) = \sum_{r=0}^{k} C_n^r \theta^r (1 - \theta)^{n-r}$$

is a continuous and monotone decreasing function in θ (for $k < n$), viz.,

$$F'\left(\frac{k}{n}; \theta\right) = -nC_{n-1}^k \theta^k (1 - \theta)^{n-k-1} < 0,\ k < n.$$

Consequently, the sought-for interval is defined by the solutions of the equations $F(T; \theta) = 1 - F(T - 0; \theta) = (1 - \gamma)/2$, which are just as given in the statement of the problem. The expression for the boundaries T_1, T_2 in terms of the quantiles of the beta distribution follows from

$$\sum_{r=k}^{n} C_n^r p^r (1 - p)^{n-r} = \frac{1}{\mathrm{B}(k, n - k + 1)} \int_0^p x^{k-1}(1 - x)^{n-k}\, dx.$$

If the number n of observations is large, then we may use the asymptotic theory of maximum likelihood estimates to find the approximate confidence interval for θ quickly. In this case the m.l.e. is $\hat{\theta}_n = \overline{X}$ (see Problem 2.84), and the information function is $i(\theta) = [\theta(1 - \theta)]^{-1}$ (see Problem 2.43). The sought-for interval has the form $(\overline{X} \pm c_\gamma\sqrt{\overline{X}(1 - \overline{X})/n})$.

2.134. For large n we have

$$\begin{aligned}\gamma &= \mathbf{P}_\theta(\sqrt{n}|\overline{X} - \theta|/\sqrt{\theta(1 - \theta)} < c_\gamma) = \mathbf{P}_\theta\left((\overline{X} - \theta)^2 < \frac{c_\gamma^2}{n}\theta(1 - \theta)\right) \\ &= \mathbf{P}_\theta\left(\theta^2\left(1 + \frac{c_\gamma^2}{n}\right) - 2\theta\left(\overline{X} + \frac{c_\gamma^2}{2n}\right) + \overline{X}^2 < 0\right) \\ &= \mathbf{P}_\theta((\theta - T_1)(\theta - T_2) < 0) = \mathbf{P}_\theta(T_1 < \theta < T_2),\end{aligned}$$

where

$$T_{1,2} = T_{1,2}(\mathbf{X}) = \left(\overline{X} + \frac{c_\gamma^2}{2n} \mp \sqrt{\overline{X}(1 - \overline{X})\frac{c_\gamma^2}{n} + \frac{c_\gamma^4}{4n^2}}\right) \Big/ \left(1 + \frac{c_\gamma^2}{n}\right).$$

Thus, (T_1, T_2) is the sought-for approximate γ-confidence interval. If we neglect the terms of the order $1/n$, then this interval is reduced to the interval $(\overline{X} \pm c_\gamma\sqrt{\overline{X}(1 - \overline{X})/n})$ obtained in the previous problem.

2.135. Since $\mathscr{L}_\theta(2\sqrt{n}(\tau(\overline{X}) - \tau(\theta))) \to \mathscr{N}(0, 1)$, for large samples we have

$$\gamma \simeq \mathbf{P}_\theta(2\sqrt{n}|\tau(\overline{X}) - \tau(\theta)| < c_\gamma) = \mathbf{P}_\theta\left(\tau(\overline{X}) - \frac{c_\gamma}{2\sqrt{n}} < \tau(\theta) < \tau(\overline{X}) + \frac{c_\gamma}{2\sqrt{n}}\right),$$

which is equivalent to our assertion.

2.137. The solution is similar to that of Problem 2.133. Here $\mathscr{L}(nT) = \Pi(n\theta)$, and T assumes the values of k/n, $k = 0, 1, 2, \ldots$. Its distribution function

$$F\left(\frac{k}{n}; \theta\right) = \sum_{r=0}^{k} e^{-n\theta}\frac{(n\theta)^r}{r!}$$

is continuous and monotone decreasing in θ, i.e.,

$$F'\left(\frac{k}{n}; \theta\right) = -ne^{-n\theta}\frac{(n\theta)^k}{k!} < 0.$$

Therefore, the sought-for interval is defined by the solutions of the equations $F(T; \theta) = 1 - F(T - 0; \theta) = (1 - \gamma)/2$ which have the indicated form. The expression for the boundaries T_1, T_2 in terms of the $\chi^2_{p,k}$-quantiles follows from

$$\mathbf{P}(\chi^2_{2k} > 2\lambda) = \sum_{r=0}^{k-1} e^{-\lambda}\frac{\lambda^r}{r!},$$

where $\mathscr{L}(\chi^2_{2k}) = \chi^2_{(2k)}$. This can easily be verified.

For large n (by Problems 2.84 and 2.43) we have the approximation

$$\mathscr{L}_\theta\left(\sqrt{\frac{n}{\overline{X}}}(\overline{X} - \theta)\right) \to \mathscr{N}(0, 1)$$

for $\hat{\theta}_n = \overline{X}$, whence we find the approximation interval.

2.138. In the first case we have

$$\begin{aligned}\gamma &\simeq \mathbf{P}_\theta(2\sqrt{n}|\sqrt{\overline{X}} - \sqrt{\theta}| < c_\gamma)\\ &= \mathbf{P}_\theta\left(\max\left(0, \sqrt{\overline{X}} - \frac{c_\gamma}{2\sqrt{n}}\right) < \sqrt{\theta} < \sqrt{\overline{X}} + \frac{c_\gamma}{2\sqrt{n}}\right).\end{aligned}$$

Whence follows that

$$\left(\left[\max\left(0, \sqrt{\overline{X}} - \frac{c_\gamma}{2\sqrt{n}}\right)\right]^2, \left(\sqrt{\overline{X}} + \frac{c_\gamma}{2\sqrt{n}}\right)^2\right)$$

is the sought-for approximate γ-confidence interval for θ. Using another approximation, we obtain

$$\gamma \simeq \mathbf{P}_\theta(\sqrt{n}|\overline{X} - \theta|/\sqrt{\theta} < c_\gamma) = \mathbf{P}_\theta((\overline{X} - \theta)^2 < c_\gamma^2\theta/n)$$
$$= \mathbf{P}_\theta\left(\theta^2 - 2\theta\left(\overline{X} + \frac{c_\gamma^2}{2n}\right) + \overline{X}^2 < 0\right) = \mathbf{P}_\theta(T_1 < \theta < T_2),$$

where $T_{1,2} = T_{1,2}(\mathbf{X}) = \overline{X} + \dfrac{c_\gamma^2}{2n} \mp \sqrt{\overline{X}\dfrac{c_\gamma^2}{n} + \dfrac{c_\gamma^4}{4n^2}}$.

Thus, (T_1, T_2) is also a γ-confidence interval for θ.

These two intervals are equivalent to the interval $(\overline{X} \pm c_\gamma\sqrt{\overline{X}/n})$ of the previous problem up to the terms of order $1/n$.

2.140. Using the result of Problem 2.96, we find the sought-for interval $(\hat{\theta}_n \pm c_\gamma\sqrt{\hat{\theta}_n/n\mu'(\hat{\theta}_n)})$, where $\hat{\theta}_n$ is the solution of the equation $\mu(\theta) = \overline{X}$, and $\mu(\theta)$ is the theoretical mean of the distribution. For the distribution $\overline{Bi}(r, \theta)$ the interval is

$$\left(\frac{\overline{X}}{r + \overline{X}} \pm c_\gamma\sqrt{\frac{r\overline{X}}{n(r + \overline{X})^3}}\right).$$

2.141. The sought-for interval is $(\hat{\theta}_n \pm c_\gamma/\sqrt{ni(\hat{\theta}_n)})$, where $i(\theta) = \lambda/\theta^2$, $\hat{\theta}_n = \overline{X}/\lambda$. As a result, we find the interval $\overline{X}\lambda^{-1}(1 \pm c_\gamma/\sqrt{\lambda n})$.

If we use the approximation $\mathscr{L}_\theta(\sqrt{\lambda n}(\ln \hat{\theta}_n - \ln \theta)) \sim \mathscr{N}(0, 1)$ (see the solution to Problem 2.109), then we will have

$$\gamma = \mathbf{P}_\theta(\sqrt{\lambda n}|\ln \hat{\theta}_n - \ln \theta| < c_\gamma) = \mathbf{P}_\theta\left(\ln \hat{\theta}_n - \frac{c_\gamma}{\sqrt{\lambda n}} < \ln \theta < \ln \hat{\theta}_n + \frac{c_\gamma}{\sqrt{\lambda n}}\right)$$
$$= \mathbf{P}_\theta\left(\hat{\theta}_n \exp\left\{-\frac{c_\gamma}{\sqrt{\lambda n}}\right\} < \theta < \hat{\theta}_n \exp\left\{\frac{c_\gamma}{\sqrt{\lambda n}}\right\}\right),$$

i.e., $(\overline{X}\lambda^{-1}e^{-c_\gamma/\sqrt{\lambda n}}, \overline{X}\lambda^{-1}e^{c_\gamma/\sqrt{\lambda n}})$ is another approximation of the γ-confidence interval which coincides with the first one up to the terms of order $1/n$.

2.142. By Problem 2.109 we have

$$\gamma \simeq \mathbf{P}_\theta(\sqrt{2n}|\ln \hat{\theta}_n - \ln \theta| < c_\gamma)$$
$$= \mathbf{P}_\theta\left(\hat{\theta}_n \exp\left\{-\frac{c_\gamma}{\sqrt{2n}}\right\} < \theta < \hat{\theta}_n \exp\left\{\frac{c_\gamma}{\sqrt{2n}}\right\}\right),$$

where $\hat{\theta}_n = \left[\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2\right]^{1/2}$. This interval is reduced to $\hat{\theta}_n(1 \pm c_\gamma/\sqrt{2n})$ if we neglect the terms of order $1/n$ and use the standard approximation $\mathscr{L}_\theta(\hat{\theta}_n) \sim \mathscr{N}\left(\theta, \dfrac{\theta_n^2}{2n}\right)$ (see Problem 2.88).

2.143. The result of Problem 2.87 implies that the sought-for interval based on the standard approximation for m.l.e.'s has the form $(\hat{\tau}_n \pm c_\gamma \sigma_\tau(\hat{\theta}_n)/\sqrt{n})$, where $\hat{\tau}_n = \Phi\left(\frac{x_0 - \overline{X}}{S}\right)$, $\hat{\theta}_n = (\hat{\theta}_{1n}, \hat{\theta}_{2n}) = (\overline{X}, S)$, $\sigma_\tau^2(\theta)$ (see Problem 2.87).

2.144. In this case the model is defined by the $(N-1)$-dimensional parameter $\theta = (p_1, \ldots, p_{N-1})$ (see Problem 2.63) and the likelihood function of the sample $\mathbf{X} = (X_1, \ldots, X_n)$ is

$$L(\mathbf{X};\theta) = \prod_{j=1}^{N} p_j^{\nu_j}$$
$$= \exp\left\{\sum_{j=1}^{N-1} \nu_j \ln \frac{p_j}{1-p_1-\ldots-p_{N-1}} + n\ln(1-p_1-\ldots-p_{N-1})\right\},$$

where ν_j is the number of units of the sample $\mathbf{X}$, which are equal to a_j, $j = 1, \ldots, N$. We now find the solution to the likelihood equations $\frac{\partial \ln L}{\partial p_j} = 0$, $j = 1, \ldots, N-1$ (i.e., the maximum likelihood estimates for the parameters $p_1, \ldots, p_{N-1}$), which has the form $\hat{p}_j = \nu_j/n, j = 1, \ldots, N-1$. The information matrix $\mathbf{I}(\theta)$ for this model was found in Problem 2.45. By the asymptotic theory of m.l.e.'s we have $\mathscr{L}_\theta(\sqrt{n}(\hat{\theta}_n - \theta)) \to \mathscr{N}(\mathbf{0}, \mathbf{I}^{-1}(\hat{\theta}_n))$ as $n \to \infty$ and, therefore, $\mathscr{L}_\theta(Q_n(\theta)) \to \chi^2(N-1)$ as $n \to \infty$, where the quadratic form is

$$Q_n(\theta) = n(\hat{\theta}_n - \theta)'\mathbf{I}(\hat{\theta}_n)(\hat{\theta}_n - \theta) = n\sum_{r,s=1}^{N-1}(\hat{p}_r - p_r)(\hat{p}_s - p_s)/\hat{p}_N$$
$$+ n\sum_{r=1}^{N-1}(\hat{p}_r - p_r)^2/\hat{p}_r = \sum_{r=1}^{N}(\nu_r - np_r)^2/\nu_r.$$

Whence as $n \to \infty$

$$\mathbf{P}_\theta\left(\sum_{r=1}^{N}(\nu_r - np_r)^2/\nu_r < \chi^2_{\gamma,N-1}\right) \to \gamma.$$

This means that the sought-for asymptotic γ-confidence region for the parameters $p_1, \ldots, p_N$ has the form

$$\mathscr{G}_\gamma(\mathbf{X}) = \left\{(p_1, \ldots, p_N):\ \sum_{r=1}^{N}(\nu_r - np_r)^2/\nu_r < \chi^2_{\gamma,N-1},\ 0 < p_i < 1,\right.$$
$$\left. i = 1, \ldots, N,\ \sum_{i=1}^{N} p_i = 1\right\}$$

and is an intersection of the inside of the N-dimensional ellipsoid

$$\sum_{r=1}^{N} \frac{(\nu_r - np_r)^2}{\nu_r} < \chi^2_{\gamma, N-1}$$

with the hyperplane $p_1 + \ldots + p_N = 1$ within the zone $0 < p_i < 1$, $i = 1, \ldots, N$. For $N = 2$ we obtain a result similar to that of Problem 2.134.

2.145. By Fisher's theorem (see also Problem 1.56) $\overline{X}$ and S^2 are independent, whence follows that $\overline{X} - X_{n+1}$ and S^2 are also independent. But $\mathscr{L}_\theta(\overline{X} - X_{n+1}) = \mathscr{N}\left(0, \ \theta_2^2 \frac{n+1}{n}\right)$ and $\mathscr{L}_\theta\left(\frac{nS^2}{\theta_2^2}\right) = \chi^2(n-1)$. We then find Student's ratio $t_{n-1} = \sqrt{\frac{n-1}{n+1}} \frac{\overline{X} - X_{n+1}}{S}$ and

$$\gamma = \mathbf{P}_\theta\left(\sqrt{\frac{n-1}{n+1}} \frac{|\overline{X} - X_{n+1}|}{S} < t_{(1+\gamma)/2, n-1}\right)$$

$$= \mathbf{P}_\theta\left(\overline{X} - t_{(1+\gamma)/2, n-1} S \sqrt{\frac{n+1}{n-1}} < X_{n+1} < \overline{X} + t_{(1+\gamma)/2, n-1} S \sqrt{\frac{n+1}{n-1}}\right),$$

q.e.d.

2.146. For the given data we have $\bar{x} = 4.196$, $s = 0.226$, $t_{0.975,4} = 2.776$. Consequently, the required interval is (3.43, 4.96).

2.147. For large n we have

$$\gamma \simeq \mathbf{P}\left(\frac{|\xi - n|}{\sqrt{2n}} \leqslant c_\gamma\right) = \mathbf{P}((\xi - n)^2 \leqslant 2nc_\gamma^2)$$

$$= \mathbf{P}(n^2 - 2n(\xi + c_\gamma^2) + \xi^2 \leqslant 0) = \mathbf{P}(n_1 \leqslant n \leqslant n_2),$$

where $n_{1,2} = n_{1,2}(\xi) = \xi + c_\gamma^2 \mp c_\gamma\sqrt{2\xi + c_\gamma^2}$.

Since $c_{0.9} = 1.645$, the sought-for interval is (131, 189).

TO CHAPTER 3

3.1. We have two groups with frequencies $h_1 = 2048$ and $h_2 = n - h_1 = 1992$. Here the null hypothesis is H_0: $p = q = 1/2$, and the expected frequencies are $np = nq = n/2 = 2020$. The test statistic is $X_2^2 = \sum_{i=1}^{2} \frac{(h_i - np_i)^2}{np_i} = 0.776$.

For large n this quantity is approximately distributed as χ^2 with one degree of freedom. From the table of quantiles for the χ^2-distribution we find $\chi^2_{0.95,1} = 3.84$, $\chi^2_{0.9,1} = 2.71$. Since 0.776 lies within the boundaries, we conclude that the data are compatible with the hypothesis H_0.

3.3. The test statistic $X_n^2 = \sum_{i=1}^{3} \frac{(h_i - np_i)^2}{np_i} = 11.13$ is compared with the critical boundary $\chi^2_{0.95,2} = 5.99$. Since $11.13 > 5.99$, the hypothesis H_0 is rejected.

3.7. The expected number of readings in every interval is $np_i = 500/12 = 41.67$. The test statistic is $X_n^2 = 10.00$, which is smaller than the critical boundary $\chi^2_{0.9,11} = 17.3$, i.e., the agreement is good. The hypothesis H_0 is not rejected for the significance level $\alpha < 0.55$.

3.8. The test statistic is $X_n^2 = 0.47$, $\chi^2_{0.9,3} = 6.25$, i.e., the agreement at $\alpha \leqslant 0.9$ is good.

3.10. The boundaries of the intervals are sought from the equations $1 - e^{-x_1} = 1/4$, $e^{-x_j} - e^{-x_{j+1}} = 1/4$, $j = 1, 2$. We have $x_1 = 0.288$, $x_2 = 0.693$, $x_3 = 1.386$. Grouping within these intervals gives the frequency vector $h = (9, 9, 17, 15)$. The test statistic $X_{50}^2 = \sum_{j=1}^{4} \frac{(h_j - np_j)^2}{np_j} = 4.08$ is smaller than the critical boundary $\chi^2_{0.9,3} = 6.25$, i.e., the hypothesis H_0 is not rejected.

3.11. Since $\mathbf{P}(\xi \leqslant x|H_0) = 1 - e^{-x/\theta}$, the probabilities $p_j(\theta) = \mathbf{P}(\xi \in \Delta_j|H_0)$ here become

$$p_j(\theta) = e^{-(j-1)a/\theta}(1 - e^{-a/\theta}),\ j = 1, \ldots, N-1,\ p_N(\theta) = e^{-(N-1)a/\theta},$$

and the equation for finding the multinomial m.l.e. $\hat{\theta}_n$ [7, p. 150] has the form

$$\sum_{j=1}^{N} h_j p_j'(\theta)/p_j(\theta) = \sum_{j=1}^{N-1} h_j(j - 1 - je^{-a/\theta})/(1 - e^{-a/\theta}) + (N-1)h_N = 0.$$

We write $z = e^{-a/\theta}$ and find

$$z\left(\sum_{j=1}^{N} jh_j - h_N\right) = \sum_{j=1}^{N} jh_j - n.$$

Consequently, the m.l.e. is

$$\hat{z}_n = \left(\sum_{j=1}^{N} jh_j - n\right) \Big/ \left(\sum_{j=1}^{N} jh_j - h_N\right),$$

and the respective estimates for the probabilities $p_j(\theta)$ have the form

$$\hat{p}_j = \hat{z}_N^{j-1}(1 - \hat{z}_N),\ j = 1, \ldots, N-1,\ \hat{p}_N = \hat{z}_N^{N-1}.$$

According to the general theory [7, pp. 149-150], if the conditions $h_j \geqslant 5$, $j = 1, \ldots, N$, hold for large n, then the respective χ^2 goodness of fit test rejects the hypothesis H_0 if and only if

$$\hat{X}_n^2 = \sum_{j=1}^{N} (h_j - n\hat{p}_j)^2/(n\hat{p}_j) \geqslant \chi^2_{1-\alpha,N-2},$$

where α is the chosen significance level.

For the data of Problem 1.21 and the given choice of the parameters N and a we have $h_1 = 28$, $h_2 = 16$, $h_3 = 6$, $\hat{\theta}_n = \dfrac{h_2 + 2h_3}{2n - h_1} = \dfrac{7}{18}$, and $\hat{X}_n^2 = 1.96\ldots$. Since $\chi^2_{0.9,1} = 2.71$, the hypothesis H_0 is verified by the data at the significance level $\alpha \leqslant 0.1$.

3.12. We are dealing with a polynomial model with $N = 4$ outcomes and probabilities $p_1, \ldots, p_4$ which under the hypothesis H_0 have the indicated form, i.e., are functions of one unknown parameter. In order to estimate θ, we must solve the equation $\sum_{j=1}^{4} h_j p_j'(\theta)/p_j(\theta) = 0$, which in our case has the form

$$\frac{h_1}{2 + \theta} - \frac{h_2 + h_3}{1 - \theta} + \frac{h_4}{\theta} = 0.$$

This equation is reduced to

$$\varphi(\theta) \equiv n\theta^2 + (h_4 + 2h_2 + 2h_3 - h_1)\theta - 2h_4 = 0$$

since $\varphi(0) = -2h_4 < 0$, $\varphi(1) = 3(h_2 + h_3) > 0$, the latter equation has the only root $\hat{\theta}_n$ in the interval $(0, 1)$. Consequently, the χ^2 goodness of fit test rejects the hypothesis H_0 at the significance level α only in the case of

$$\sum_{j=1}^{4} h_j^2/(np_j(\hat{\theta}_n)) - n \geqslant \chi^2_{1-\alpha,2},$$

or

$$h_1^2/(n(2 + \hat{\theta}_n)) + (h_2^2 + h_3^2)/(n(1 - \hat{\theta}_n)) + h_4^2/(n\hat{\theta}_n) \geqslant (\chi^2_{1-\alpha,2} + n)/4.$$

3.14. We use the χ^2 goodness of fit test. The estimate for θ is $\hat{\theta} = \sum ih_i = 3.870$. We calculate the estimates for the probabilities $\hat{p}_i = e^{-\hat{\theta}} \dfrac{\hat{\theta}^i}{i!}$, $i = 0, 1, \ldots, 10$, and the value $\hat{X}_n^2 = \displaystyle\sum_{i=0}^{10} \frac{(h_i - n\hat{p}_i)^2}{n\hat{p}_i} = 13.05$. We have $k = 9$ degrees of freedom. Since $\chi^2_{0.95,9} = 16.9 > 13.05$, the hypothesis H_0 is not rejected.

3.15. Here $\hat{\theta} = 1.54$, $\hat{X}_n^2 = 7.95$, $k = 6$, $\chi^2_{0.9,6} = 10.6$. The data fit the model.

3.16. Here $\hat{\theta} = 0.928$, $\hat{p}_i = e^{-\hat{\theta}} \dfrac{\hat{\theta}^i}{i!}$, $i = 0, 1, \ldots, 5$, $\hat{X}_n^2 = 2.172$, $k = 6$, $\chi^2_{0.95,4} = 9.49$. The data fit well.

3.17. For $\mathscr{L}(\xi) = Bi(2, \theta)$ the probabilities of the outcomes are

$$p_1(\theta) = \mathbf{P}(\xi = 0) = (1 - \theta)^2,$$

$$p_2(\theta) = \mathbf{P}(\xi = 1) = 2\theta(1 - \theta),$$

$$p_3(\theta) = \mathbf{P}(\xi = 2) = \theta^2.$$

We find the estimates for the parameter θ from the equation

$$\sum_{j=1}^{3} h_j p_j'(\theta)/p_j(\theta) = -\frac{2h_1}{1-\theta} + \frac{1-2\theta}{\theta(1-\theta)} h_2 + \frac{2h_3}{\theta} = 0.$$

The estimate is $\hat{\theta}_n = (h_2 + 2h_3)/2n$.

Here $h_1 = 476$, $h_2 = 1017$, $h_3 = 527$, $n = 2020$. Therefore, $\hat{\theta}_n = 0.513$. We have

$$\hat{X}_n^2 = \sum_{j=1}^{3} (h_j - np_j(\hat{\theta}_n))^2/(np_j(\hat{\theta}_n)) = 0.116$$

and compare the result with $\chi^2_{1-\alpha,1}$. Since $\chi^2_{0.3,1} = 0.148$, the hypothesis is accepted for any significance level $\alpha \leqslant 0.7$.

3.19. We substitute $\mathbf{p} = \mathbf{p}^{(n)} = (p_1^{(n)}, \ldots, p_N^{(n)})$ into the formula

$$\mathbf{E}(X_n^2|\mathbf{p}) = n \sum_{j=1}^{N} (p_j - p_j^0)^2/p_j^0 + \sum_{j=1}^{N} p_j(1 - p_j)/p_j^0$$

and take into account that $\sum_{j=1}^{N} p_j^0 = 1$, $\sum_{j=1}^{N} \beta_j = 0$. We then find

$$\mathbf{E}(X_n^2|\mathbf{p}^{(n)}) = N - 1 + \sum_{j=1}^{N} \beta_j^2/p_j^0 + O(1/\sqrt{n}).$$

Since

$$R_{ks} = \sum_{j=1}^{N} p_j^{(n)^k}/p_j^{0^s} = \sum_{j=1}^{N} p_j^{0^{k-s}}\left(1 + \frac{\beta_j}{\sqrt{n}p_j^0}\right)^k,$$

we have

$$R_{32} = 1 + \frac{3}{n} \sum_{j=1}^{N} \beta_j^2/p_j^0 + O(n^{-3/2}),$$

$$R_{21} = 1 + \frac{1}{n} \sum_{j=1}^{N} \beta_j^2/p_j^0,$$

$$R_{22} = N + O(1/\sqrt{n}), \quad R_{11} = N + O(1/\sqrt{n}), \quad R_{12} = O(1).$$

Using the formula

$$\mathbf{D}(X_n^2|\mathbf{p}) = 4\,\frac{(n-1)(n-2)}{n}(R_{32} - R_{21}^2) + 2\,\frac{n-1}{n}(3R_{22} - 2R_{21}R_{11} - R_{21}^2) + \frac{1}{n}(R_{12} - R_{11}^2),$$

we find

$$\mathbf{D}(X_n^2|\mathbf{p}^{(n)}) = 2(N-1) + 4\sum_{j=1}^{N} \beta_j^2/p_j^0 + O(1/\sqrt{n}).$$

3.20. We have

$$\begin{aligned}(1-p_j^{(n)})^n &= \exp[n\ln(1-p_j^{(n)})] = \exp[-np_j^{(n)} + O(n^{-1})]\\ &= \exp[-\varrho - \varrho b_j/n^{1/4} + O(n^{-1})]\\ &= e^{-\varrho}\{1 - \varrho b_j/n^{1/4} + \varrho^2 b_j^2/2\sqrt{n} + O(n^{-3/4})\}.\end{aligned}$$

By substituting this expansion into $\mathbf{E}(\mu_0|H_1^{(n)}) = \sum_{j=1}^{N}(1-p_j^{(n)})^n$, we arrive at the sought-for expression for the mean.

We use the formula

$$\mathbf{D}\mu_0 = 2\sum_{i<j}[(1-p_i-p_j)^n - (1-p_i)^n(1-p_j)^n] - \sum_{j=1}^{N}(1-p_j)^{2n} + \mathbf{E}\mu_0$$

to investigate the variance. The asymptotics of the second sum under the hypothesis $H_1^{(n)}$ is found as above and is equal to $Ne^{-2\varrho}$ up to the main term.

Let us estimate the general term of the first sum. We have

$$\begin{aligned}(1-p_i-p_j)^n - (1-p_i)^n(1-p_j)^n &= (1-p_i-p_j)^n - (1-p_i-p_j+p_ip_j)^n\\ &= \exp\left\{-n(p_i+p_j) - \frac{n}{2}(p_i+p_j)^2 + O(N^{-2})\right\}\\ &\quad - \exp\left\{-n(p_i+p_j) + np_ip_j - \frac{n}{2}(p_i+p_j)^2 + O(N^{-2})\right\}\\ &= \exp\left\{-n(p_i+p_j) - \frac{n}{2}(p_i+p_j)^2\right\}\\ &\quad \times [\exp\{O(N^{-2})\} - \exp\{np_ip_j + O(N^{-2})\}]\\ &= -np_ip_je^{-np_i-np_j} + O(N^{-2}).\end{aligned}$$

The first sum can be represented as

$$\begin{aligned}&-2n\sum_{i<j} p_ip_je^{-np_i-np_j} + O(1)\\ &= -n\left(\sum_{j=1}^{N} p_je^{-np_j}\right)^2 + n\sum_{j=1}^{N} p_j^2e^{-2np_j} + O(1).\end{aligned}$$

Here the second term is $O(1)$, while $\sum_{j=1}^{N} p_je^{-np_j} = e^{-\varrho} + o(1)$. Hence, the entire expression is equal to $-N\varrho e^{-2\varrho} + O(1)$. Then

$$\mathbf{D}(\mu_0|N_1^{(n)}) = -N\varrho e^{-2\varrho} - Ne^{-2\varrho} + Ne^{-\varrho} + O(N^{1/2}).$$

3.21. We apply the χ^2-test for uniformity. The test statistic is

$$X_n^2 = n_1 n_2 \sum_{i=1}^{s} \frac{1}{\nu_{i1} + \nu_{i2}} (\nu_{i1}/n_1 - \nu_{i2}/n_2)^2 = 2.18,\ s = 4,\ k = 2,$$

and the critical boundary of the test is $\chi^2_{1-\alpha,(s-1)(k-1)} = \chi^2_{0.9,3} = 6.25$. Then $X_n^2 < \chi^2_{0.9,3}$, and the agreement is good.

3.23. (1) Consider the difference $\Delta_{ij} = \nu_{ij} - \nu_{i\cdot}\nu_{\cdot j}/n$, $i, j = 1, 2$. We can directly check that all the sums $\Delta_{i1} + \Delta_{i2}$ and $\Delta_{1j} + \Delta_{2j}$ are zero. For example,

$$\Delta_{11} + \Delta_{12} = \nu_{11} + \nu_{12} - \frac{\nu_{1\cdot}\nu_{\cdot 1}}{n} - \frac{\nu_{1\cdot}\nu_{\cdot 2}}{n}$$

$$= \nu_{1\cdot} - \frac{\nu_{1\cdot}}{n}(\nu_{\cdot 1} + \nu_{\cdot 2}) = \nu_{1\cdot} - \nu_{1\cdot} = 0.$$

Thus, the absolute values of the four quantities Δ_{ij} are equal and, therefore,

$$X_n^2 = n \sum_{i,j=1}^{2} \frac{\Delta_{ij}^2}{\nu_{i\cdot}\nu_{\cdot j}} = n\Delta_{11}^2 \sum_{i,j=1}^{2} \frac{1}{\nu_{i\cdot}\nu_{\cdot j}} = \frac{n^3\Delta_{11}^2}{\nu_{1\cdot}\nu_{2\cdot}\nu_{\cdot 1}\nu_{\cdot 2}}.$$

We then have

$$n\Delta_{11} = \nu_{\cdot 1}\left[\frac{\nu_{11}}{\nu_{\cdot 1}}(\nu_{\cdot 1} + \nu_{\cdot 2}) - (\nu_{11} + \nu_{12})\right] = \nu_{\cdot 1}\nu_{\cdot 2}\left(\frac{\nu_{11}}{\nu_{\cdot 1}} - \frac{\nu_{12}}{\nu_{\cdot 2}}\right)$$

and, finally,

$$X_n^2 = \frac{n\nu_{\cdot 1}\nu_{\cdot 2}}{\nu_{1\cdot}\nu_{2\cdot}}\left(\frac{\nu_{11}}{\nu_{\cdot 1}} - \frac{\nu_{12}}{\nu_{\cdot 2}}\right)^2 = Z_n^2.$$

(2) Note that the random variables ν_{11} and ν_{12} are independent by the statement of the problem, and for some $p \in (0, 1)$, $j = 1, 2$, we have $\mathscr{L}(\nu_{1j}) = Bi(n_j, p)$ if the hypothesis H_0 is true. For $n_1, n_2 \to \infty$ we find by the De Moivre-Laplace theorem that $\mathscr{L}(\nu_{1j}) \sim \mathscr{N}(n_j p, n_j pq)$, $q = 1 - p$, or $\mathscr{L}(\nu_{1j}/n_j) \sim \mathscr{N}\left(p, \frac{pq}{n_j}\right)$, $j = 1, 2$. Then

$$\mathscr{L}\left(\frac{\nu_{11}}{n_1} - \frac{\nu_{12}}{n_2}\right) \sim \mathscr{N}\left(0, \frac{npq}{n_1 n_2}\right).$$

Thus, under the hypothesis H_0 the random variable

$$\zeta_{n_1 n_2} = \left(\frac{\nu_{11}}{n_1} - \frac{\nu_{12}}{n_2}\right)\sqrt{\frac{n_1 n_2}{npq}}$$

is asymptotically normal as $\mathcal{N}(0, 1)$. Since

$$Z_n = \zeta_{n_1 n_2}\sqrt{\frac{n^2 pq}{\nu_{1\cdot}\nu_{2\cdot}}},$$

$\frac{\nu_{\cdot 1}}{n} \xrightarrow{\mathbf{P}} p$ under H_0, we have $\sqrt{\frac{n^2pq}{\nu_{1\cdot}\nu_{2\cdot}}} \xrightarrow{\mathbf{P}} 1$. It follows that the limiting distributions of the random variables Z_n and $\zeta_{n_1 n_2}$ coincide for the hypothesis H_0, q.e.d.

Finally, under the given alternative the mean value of the difference $\frac{\nu_{11}}{n_1} - \frac{\nu_{12}}{n_2}$ is $p_1 - p_2 > 0$. Therefore, while testing the hypothesis H_0 against the alternative H_1, the critical region should be chosen as $\{Z_n > t_\alpha\}$. Since $\mathbf{P}(Z_n > t_\alpha|H_0) = \Phi(-t_\alpha)$, at the significance level α the critical boundary has the form

$$t_\alpha = -\Phi^{-1}(\alpha) = \Phi^{-1}(1 - \alpha) = u_{1-\alpha}.$$

Remark. If for any hypothesis defined by the probabilities p_1, p_2 as $n_1, n_2 \to \infty$, we have

$$\mathcal{L}\left(\frac{\nu_{11}}{n_1} - \frac{\nu_{12}}{n_2}\right) \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}\right),$$

then, by reasoning in the same way, we may show that for the close alternative $H_1^{(n)}$: $(p_1 - p_2)/\sqrt{p_1q_1} = a/\sqrt{n}$, $a \neq 0$,

$$\mathcal{L}(Z_n|H_1^{(n)}) \to \mathcal{N}(a\sqrt{\gamma(1-\gamma)}, 1), \ \gamma = \lim_{n\to\infty}\frac{n_1}{n}.$$

This allows us to compute the power test for such alternatives.

3.24. Since (see Problem 1.54) we have

$$\mathcal{L}(\nu_1, \ldots, \nu_N|\nu_1 + \ldots + \nu_N = n) = M(n; p_1, \ldots, p_N)$$

for $p_j = \theta_j \Big/ \sum_{i=1}^{N} \theta_i$, $j = 1, \ldots, N$, the hypothesis H_0 is equivalent here to the hypothesis about equal probabilities of outcomes in a polynomial model. Consequently, the χ^2 goodness of fit test has the form

$$X_n^2 = \frac{N}{n}\sum_{j=1}^{N}\left(\nu_j - \frac{n}{N}\right)^2 \geq \chi^2_{1-\alpha, N-1}.$$

Remark. We use $\bar{\nu}$, $S^2 = \frac{1}{N}\sum_{j=1}^{N}(\nu_j - \bar{\nu})^2$ to denote the sample mean and variance and write the statistic X_n^2 as $X_n^2 = NS^2/\bar{\nu}$. Under the hypothesis

H_0 the theoretical mean and variance of the observations are equal and, therefore, if H_0 is true, then we have $S^2/\bar{\nu} \xrightarrow{\mathbf{P}} 1$ as $n \to \infty$ or $X_n^2 \xrightarrow{\mathbf{P}} N$.

3.25. (1) Since the joint distribution density of the order statistics $X_{(1)}, \ldots, X_{(n)}$ is

$$g(x_1, \ldots, x_n) = n!, \; 0 \leqslant x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n \leqslant 1$$

(see Problem 1.31), by the formula for the total probability, the unconditional distribution of the vector $\varkappa = (\varkappa_1, \ldots, \varkappa_{n+1})$ has the form

$$\begin{aligned}
&\mathbf{P}(\varkappa_i = k_i, \; i = 1, \ldots, n+1) \\
&= \frac{m!}{k_1! \ldots k_{n+1}!} \int \prod_{j=1}^{n+1} (x_j - x_{j-1})^{k_j} g(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\
&= \frac{m!\,n!}{k_1! \ldots k_{n+1}!} \int_0^1 dx_1 \int_{x_1}^1 dx_2 \ldots \int_{x_{n-1}}^1 \prod_{j=1}^{n+1} (x_j - x_{j-1})^{k_j}\, dx_n
\end{aligned}$$

(here $k_1 + \ldots + k_{n+1} = m$, $x_0 = 0$, $x_{n+1} = 1$). Integrating this with respect to x_n, x_{n-1}, etc., and applying the formula

$$\int_a^1 (x - a)^r (1 - x)^s\, dx = \frac{r!\,s!}{(r + s + 1)!} (1 - a)^{r+s+1},$$

we get

$$\begin{aligned}
\mathbf{P}(\varkappa_i = k_i, \; i = 1, \ldots, n+1) &= \frac{m!\,n!}{k_1! \ldots k_{n+1}!} \frac{k_n!\,k_{n+1}!}{(k_n + k_{n+1} + 1)!} \\
&\times \frac{k_{n-1}!(k_n + k_{n+1} + 1)!}{(k_{n-1} + k_n + k_{n+1} + 2)!} \ldots \frac{k_1!(k_2 + \ldots + k_{n+1} + n - 1)!}{(k_1 + k_2 + \ldots + k_{n+1} + n)!} \\
&= \frac{m!\,n!}{(m + n)!} = (C_{m+n}^n)^{-1}.
\end{aligned}$$

On the other hand, for $k_1 + \ldots + k_{n+1} = m$ and arbitrary $p \in (0, 1)$, $q = 1 - p$, we have

$$\begin{aligned}
&\mathbf{P}(\xi_i = k_i, \; i = 1, \ldots, n + 1 \mid \xi_1 + \ldots + \xi_{n+1} = m) \\
&= \frac{\mathbf{P}(\xi_i = k_i, \; i = 1, \ldots, n + 1)}{\mathbf{P}(\xi_1 + \ldots + \xi_{n+1} = m)} = \frac{\prod_{i=1}^{n+1} (p^{k_i} q)}{C_{n+m}^m p^m q^{n+1}} = (C_{n+m}^n)^{-1}
\end{aligned}$$

because $\mathcal{L}(\xi_1 + \ldots + \xi_{n+1}) = \overline{Bi}(n + 1, p)$ (see Problem 1.39 (5)). Thus, if the homogeneity hypothesis H_0 is true, then

$$\mathcal{L}(\varkappa) = \mathcal{L}(\xi_1, \ldots, \xi_{n+1} | \xi_1 + \ldots + \xi_{n+1} = m).$$

(2) We calculate

$$\mathbf{P}(s_0(n, m) = k) = \mathbf{P}\left(\sum_{i=1}^{n+1} I(\xi_i = 0) = k | \xi_1 + \ldots + \xi_{n+1} = m \right)$$
$$= \mathbf{P}\left(\sum_{i=1}^{n+1} I(\xi_i = 0) = k,\ \xi_1 + \ldots + \xi_{n+1} = m \right) \Big/ \mathbf{P}(\xi_1 + \ldots + \xi_{n+1} = m).$$

In order to find the numerator, we write the numbers of the zero values. This can be done in C_{n+1}^k different ways. We write

$$C_{n+1}^k \mathbf{P}(\xi_i > 0,\ i = 1, \ldots, n + 1 - k,\ \xi_j = 0,\ j = n + 2 - k, \ldots, n + 1)$$
$$\times \mathbf{P}(\xi_1 + \ldots + \xi_{n+1} = m | \xi_i > 0,\ i = 1, \ldots, n + 1 - k,\ \xi_j = 0,$$
$$j = n + 2 - k, \ldots, n + 1) = C_{n+1}^k p^{n+1-k} q^k \mathbf{P}(\tilde{\xi}_1 + \ldots + \tilde{\xi}_{n+1-k} = m)$$

(see (2) in the hint). For $r = 1, 2, \ldots$, we have

$$\mathbf{P}(\tilde{\xi}_1 = r) = \mathbf{P}(\xi_1 = r)/\mathbf{P}(\xi_1 > 0) = \frac{p^r q}{p} = p^{r-1} q,$$

i.e., $\mathcal{L}(\tilde{\xi}_1) = \mathcal{L}(\xi_1 + 1)$. Then $\mathcal{L}(\tilde{\xi}_1 + \ldots + \tilde{\xi}_s) = \mathcal{L}(\xi_1 + \ldots + \xi_s + s)$ and, therefore,

$$\mathbf{P}(\tilde{\xi}_1 + \ldots + \tilde{\xi}_s = m) = \mathbf{P}(\xi_1 + \ldots + \xi_s = m - s) = C_{m-1}^{s-1} q^s p^{m-s}.$$

We finally obtain

$$\mathbf{P}(s_0(n, m) = k) = C_{n+1}^k q^k p^{n+1-k} C_{m-1}^{n-k} q^{n+1-k} p^{m+k-n-1} / C_{n+m}^n p^m q^{n+1}$$
$$= C_{n+1}^k C_{m-1}^{n-k} / C_{n+m}^n$$

and

$$\mathcal{L}(s_0(n, m)) = H(n + 1, n + m, n).$$

Using the formulas for the moments of the hypergeometric distribution, we find that under the hypothesis H_0

$$\mathbf{E}s_0(n, m) = \frac{n(n + 1)}{n + m},\ \mathbf{D}s_0(n, m) = \frac{m(m - 1)n(n + 1)}{(n + m)^2(n + m - 1)}.$$

If $n, m \to \infty$ so that $m/n = \varrho > 0$, then

$$\mathbf{E}s_0(n, m) = \frac{n}{1 + \varrho} + O(1),\ \mathbf{D}s_0(n, m) = \frac{n\varrho^2}{(1 + \varrho)^3} + O(1).$$

(3) The formula for $\mathbf{P}(s_0(n, m) = k)$ given in the hint follows from (2). Let $k = (n + 1)p + t\sqrt{npq}$, $|t| \leqslant c < \infty$. Then

$$b(k;\ n + 1,\ p) = \frac{1 + o(1)}{\sqrt{2\pi npq}}\,e^{-t^2/2}$$

and $n - k = (m - 1)p - \frac{t}{\sqrt{\varrho}}\sqrt{npq}$. Therefore,

$$b(n - k;\ m - 1,\ p) = \frac{1 + o(1)}{\sqrt{2\pi mpq}}\,e^{-t^2/(2\varrho)}.$$

Since $n = (n + m)p$, we have

$$b(n;\ n + m,\ p) = \frac{1 + o(1)}{\sqrt{2\pi(n + m)pq}}$$

and unite these estimates to obtain

$$\mathbf{P}(s_0(n, m) = k) = \sqrt{\frac{n + m}{2\pi mnpq}}\,e^{-t^2(1+\varrho)/(2\varrho)}(1 + o(1)).$$

This means that for $k = \frac{n + 1}{1 + \varrho} + x\sqrt{n\varrho^2/(1 + \varrho)^3}$, $|x| \leqslant c < \infty$, we will get

$$\mathbf{P}(s_0(n, m) = k) = \frac{1 + o(1)}{\sqrt{2\pi n\varrho^2/(1 + \varrho)^3}}\,e^{-x^2/2},$$

i.e., we arrive at the normal local (and hence integral) limit theorem.

(4)

$$\mathbf{E}[s_0(n, m)|H_1] = \sum_{i=1}^{n+1} \mathbf{P}(x_i = 0|H_1).$$

We first investigate the terms for $i = 2, \ldots, n$. The probability that the block B_i is empty for fixed $X_{(i-1)} = x_1 < X_{(i)} = x_2$ is $[1 - F(x_2) + F(x_1)]^m$. By the formula for the total probability we get

$$\mathbf{P}(x_i = 0|H_1)$$

$$= \frac{n!}{(i - 2)!(n - i)!}\int_0^1 dx_1 \int_{x_1}^1 [1 - F(x_2) + F(x_1)]^m x_1^{i-2}(1 - x_2)^{n-i}\,dx_2.$$

For $i = 1$ and $i = n + 1$ the probabilities will change. By dividing their expressions by $n + 1$, we will have a zero contribution to the sum and hence can neglect them. By summing the resultant expressions, we get

$$\mathbf{E}\left[\frac{s_0(n,\ m)}{n+1}\middle| H_1\right]$$
$$= \frac{n(n-1)}{n+1}\int_0^1 dx_1 \int_{x_1}^1 [1 - F(x_2) + F(x_1)]^m (1 - x_2 + x_1)^{n-2}\, dx_2 + \varepsilon_n,$$

where $\varepsilon_n < 2/(n+1)$. We change the variables $x_1 = y_1$, $x_2 = y_1 + y_2/n$ and write

$$\mathbf{E}\left[\frac{s_0(n,\ m)}{n+1}\middle| H_1\right] = \frac{n-1}{n+1}\int_0^1 dy_1 \int_0^{(1-y_1)n} \left(1 - \frac{y_2}{n}\right)^{n-2}$$
$$\times \left[1 - \frac{F\left(y_1 + \dfrac{y_2}{n}\right) - F(y_1)}{y_2/n}\,\frac{y_2}{n}\right]^m dy_2 + \varepsilon_n.$$

In the limit as n, $m \to \infty$, $m = \varrho n$, we get

$$\lim_{n\to\infty} \mathbf{E}\left[\frac{s_0(n,\ m)}{n+1}\middle| H_1\right]$$
$$= \int_0^1 dy_1 \int_0^\infty \mathrm{e}^{-y_2(1+\varrho f(y_1))}\, dy_2 = \int_0^1 \frac{dx}{1+\varrho f(x)}$$

taking into account that the function F is differentiable. Then (see the hint)

$$1 = \left(\int_0^1 g_1(x)\, g_2(x)\, dx\right)^2 \leqslant \int_0^1 (1 + \varrho f(x))\, dx \int_0^1 \frac{dx}{1+\varrho f(x)}$$
$$= (1+\varrho)\int_0^1 \frac{dx}{1+\varrho f(x)}$$

since $\int_0^1 f(x)\, dx = 1$. The equality only holds when the functions $g_1(x)$ and $g_2(x)$ are proportional. Then $1 + \varrho f(x) \equiv \text{const}$, i.e., $f(x) \equiv \text{const}$. But $f(x)$ is the density function and hence is identically unity. Thus, if the alternative is true, then the inequality will be strict, and under the alternative the statistic $s_0(n,\ m)$ will asymptotically tend to larger values than under the hypothesis H_0. Therefore, the critical region for H_0 should be chosen in the form $\{s_0(n,\ m) > c\}$. At the significance level α the critical boundary asymptotically

tends to

$$c = c_\alpha(n) = \frac{n}{1+\varrho} - \sqrt{n}\,\frac{\varrho}{(1+\varrho)^{3/2}}\,u_\alpha.$$

3.26. We have $\hat{X}_n^2 = n\left(\sum_{ij} \frac{\nu_{ij}^2}{\nu_{i\cdot}\nu_{\cdot j}} - 1\right) = 8.09 < \chi^2_{0.95,4} = 9.49$. The hypothesis is not rejected.

3.28. (1) See the solution to Problem 3.23 (1).

(2) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent observations on (ξ_1, ξ_2) and $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{Y} = (Y_1, \ldots, Y_n)$. Then the sample means are

$$\overline{X} = \frac{\nu_{1\cdot}}{n},\ \overline{Y} = \frac{\nu_{\cdot 1}}{n},$$

the sample variances are

$$S_1^2 = S^2(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{\nu_{1\cdot}}{n}\left(1 - \frac{\nu_{1\cdot}}{n}\right) = \frac{\nu_{1\cdot}\nu_{2\cdot}}{n^2},$$

$$S_2^2 = S^2(\mathbf{Y}) = \frac{\nu_{\cdot 1}\nu_{\cdot 2}}{n^2},$$

and the sample covariance is

$$S_{12} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}\,\overline{Y} = \frac{\nu_{11}}{n} - \frac{\nu_{1\cdot}\nu_{\cdot 1}}{n^2} = \frac{\Delta_{11}}{n}.$$

Then (see the solution to Problem 1.38)

$$\varrho_n = S_{12}/S_1S_2 = n\Delta_{11}/\sqrt{\nu_{1\cdot}\nu_{2\cdot}\nu_{\cdot 1}\nu_{\cdot 2}} = n^{-1/2}Z_n.$$

The theoretical correlation coefficient is

$$\varrho = \frac{\mathbf{E}(\xi_1\xi_2) - \mathbf{E}\xi_1\mathbf{E}\xi_2}{\sqrt{\mathbf{D}\xi_1\mathbf{D}\xi_2}} = \frac{\mathbf{P}(\xi_1 = 1, \xi_2 = 1) - \mathbf{P}(\xi_1 = 1)\mathbf{P}(\xi_2 = 1)}{\sqrt{\mathbf{P}(\xi_1 = 1)\mathbf{P}(\xi_1 = 0)\mathbf{P}(\xi_2 = 1)\mathbf{P}(\xi_2 = 0)}}$$
$$= \frac{\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}}.$$

But

$$\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(B)\left[\frac{\mathbf{P}(AB)}{\mathbf{P}(B)}(\mathbf{P}(B) + \mathbf{P}(\overline{B})) - (\mathbf{P}(AB) + \mathbf{P}(A\overline{B}))\right]$$
$$= \mathbf{P}(B)\mathbf{P}(\overline{B})[\mathbf{P}(AB)/\mathbf{P}(B) - \mathbf{P}(A\overline{B})/\mathbf{P}(\overline{B})],$$

whence follows the second representation for ϱ.

(3) Consider the random variable

$$\zeta_n = \sum_{i=1}^{n} (X_i - \mathbf{P}(A))(Y_i - \mathbf{P}(B))/\sqrt{n\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}.$$

Since for H_0 we have

$$\mathbf{E}(X_i - \mathbf{P}(A))(Y_i - \mathbf{P}(B)) = \mathbf{E}(X_i - \mathbf{P}(A))\mathbf{E}(Y_i - \mathbf{P}(B)) = 0,$$

$$\mathbf{D}(X_i - \mathbf{P}(A))(Y_i - \mathbf{P}(B)) = \mathbf{E}(X_i - \mathbf{P}(A))^2(Y_i - \mathbf{P}(B))^2$$

$$= \mathbf{E}(X_i - \mathbf{P}(A))^2\mathbf{E}(Y_i - \mathbf{P}(B))^2 = \mathbf{D}X_i\mathbf{D}Y_i = \mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B}),$$

ζ_n is a normalized sum of independent and similarly distributed random variables. By the Central Limit Theorem we have

$$\mathscr{L}(\zeta_n) \to \mathscr{N}(0, 1)$$

as $n \to \infty$ and

$$\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

$$= \sum_{i=1}^{n} (X_i - \mathbf{P}(A))(Y_i - \mathbf{P}(B)) - n(\overline{X} - \mathbf{P}(A))(\overline{Y} - \mathbf{P}(B)).$$

Using (2), we may write Z_n in the form

$$Z_n = n^{1/2}S_{12}/(S_1S_2) = \frac{n^{1/2}\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\overline{X}(1 - \overline{X})\overline{Y}(1 - \overline{Y})}}$$

$$= \left[\zeta_n - \frac{n^{1/2}(\overline{X} - \mathbf{P}(A))}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})}}\,\frac{\overline{Y} - \mathbf{P}(B)}{\sqrt{\mathbf{P}(B)\mathbf{P}(\overline{B})}}\right]\left[\frac{\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}{\overline{X}(1 - \overline{X})\overline{Y}(1 - \overline{Y})}\right]^{1/2}.$$

By the theorem on the asymptotic normality of sampling moments we have

$$\mathscr{L}\left(\frac{n^{1/2}(\overline{X} - \mathbf{P}(A))}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})}}\right) \to \mathscr{N}(0, 1)$$

as $n \to \infty$. According to the law of large numbers, $\overline{X} \xrightarrow{\mathbf{P}} \mathbf{P}(A)$, $\overline{Y} \xrightarrow{\mathbf{P}} \mathbf{P}(B)$. Hence,

$$\left[\frac{\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}{\overline{X}(1 - \overline{X})\,\overline{Y}(1 - \overline{Y})}\right]^{1/2} \to 1, \quad \frac{n^{1/2}(\overline{X} - \mathbf{P}(A))(\overline{Y} - \mathbf{P}(B))}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}} \xrightarrow{\mathbf{P}} 0$$

as $n \to \infty$, and therefore the limiting distributions of Z_n and ζ_n coincide under the hypothesis H_0.

We see that $\varrho = 0$ under the hypothesis H_0, while $\varrho > 0$ under the alternative H_1, whence $Z_n \xrightarrow{\mathbf{P}} +\infty$ as $n \to \infty$. Therefore, large values of the statistic

Z_n testify that the alternative is preferable. In other words, when testing the hypothesis H_0 against the alternative H_1, we should choose the critical region in the form $\{Z_n > t_\alpha\}$. Since

$$\mathbf{P}(Z_n > t_\alpha | H_0) \doteq \Phi(-t_\alpha)$$

for the chosen significance level α the critical boundary is $t_\alpha = -\Phi^{-1}(\alpha) = \Phi^{-1}(1-\alpha) = u_{1-\alpha}$.

Remark. Reasoning along the same lines, we may show that if a close alternative is defined by the conditions

$$H_1^{(n)}: \mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B) + O(n^{-1/2}),$$
$$\varrho = \varrho^{(n)} = an^{-1/2},\ a \neq 0,$$

then $\mathscr{L}(Z_n | H_1^{(n)}) \to \mathscr{N}(a, 1)$, and hence for $a > 0$ the power of the constructed test satisfies the limiting relation

$$W_n(H_1^{(n)}) \to \Phi(a + u_\alpha).$$

3.29. Here (see Problem 3.28) $Z_n = \left(\frac{97}{360} - \frac{40}{82}\right)\sqrt{\frac{360 \times 82 \times 442}{137 \times 305}} = -3.86$ and $\hat{X}_n^2 = Z_n^2 = 14.89$. Since $\chi^2_{0.9999,1} = 10.8$, the hypothesis that the features are independent must be rejected (then the probability of erroneous decision is smaller than 10^{-4}). At the same time the data testify against the hypothesis H_1 since $Z_n < 0$ (this may be interpreted, for example, as the absence of discrimination for the women entering the university).

3.30. Here $Z_n = \left(\frac{276}{749} - \frac{3}{69}\right)\sqrt{\frac{749 \times 69 \times 818}{279 \times 539}} = 5.45$ and $\hat{X}_n^2 = Z_n^2 = 29.70$. Using the data we already have (see the solution to the previous problem), we reject the hypothesis H_0. Since $Z_n > \Phi^{-1}(0.9999) = 3.72$, the data (see the solution to Problem 3.28 (3)) verify the hypothesis H_1. The probability to make an error when rejecting H_0 and accepting H_1 is smaller than 10^{-4}.

3.31. We are verifying the randomness hypothesis H_0 [7, p. 169]. In this case the number of inversions is $T_8 = 0$, and the probability to get this value under the hypothesis H_0 is $(8!)^{-1} = 0.25 \times 10^{-4}$. Consequently, the hypothesis H_0 is rejected for any reasonable significance level.

3.32. We write $T_n^* = 6\left(T_n - \frac{n(n-1)}{4}\right)n^{-3/2}$. Then the characteristic function is

$$\mathbf{E}e^{itT_n^*} = \exp\left\{-\frac{3it(n-1)}{2\sqrt{n}}\right\}\Phi_n(e^{6itn^{-3/2}})$$
$$= \exp\left\{-\frac{3it(n-1)}{2\sqrt{n}}\right\}\prod_{r=2}^{n}\left[1 + \frac{1}{r}\sum_{k=1}^{r-1}(e^{6itkn^{-3/2}} - 1)\right].$$

For $|t| \leqslant c < \infty$, $n \to \infty$ and any $k = 1, \ldots, n$ we have

$$\exp\left(6itkn^{-3/2}\right) - 1 = 6itkn^{-3/2} - 18t^2k^2n^{-3} + O(k^3n^{-9/2}).$$

Then

$$\frac{1}{r}\sum_{k=1}^{r-1}\left(e^{6itkn^{-3/2}} - 1\right) = \frac{3it(r-1)}{n^{3/2}} - \frac{3t^2(r-1)(2r-1)}{n^3} + O\left(\frac{r^3}{n^{9/2}}\right).$$

We go to logarithms and use the formula

$$\ln(1 + x) = x - \frac{x^2}{2} + O(x^3),\ x \to 0,$$

to obtain

$$\ln \mathrm{E}e^{itT_n^*} = -\frac{3it(n-1)}{2\sqrt{n}} + \frac{3it}{n^{3/2}}\sum_{r=2}^{n}(r-1) - \frac{3t^2}{n^3}\sum_{r=2}^{n}(r-1)(2r-1)$$
$$+ \frac{9t^2}{2n^3}\sum_{r=2}^{n}(r-1)^2 + \sum_{r=2}^{n} O\left(\frac{r^3}{n^{9/2}}\right) = -\frac{t^2}{2} + O\left(\frac{1}{\sqrt{n}}\right),$$

whence follows the required assertion.

3.39. Here the likelihood ratio statistic is

$$l(\mathbf{X}) = \frac{L(\mathbf{X};\ \theta_1)}{L(\mathbf{X};\ \theta_0)} = \prod_{i=1}^{n} C_k^{X_i}\theta_1^{X_i}(1-\theta_1)^{k-X_i}\Big/\prod_{i=1}^{n} C_k^{X_i}\theta_0^{X_i}(1-\theta_0)^{k-X_i}$$
$$= \left[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right]^T\left(\frac{1-\theta_1}{1-\theta_0}\right)^{kn},\ T = \sum_{i=1}^{n} X_i,$$

and $\dfrac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} > 1$ for $\theta_1 > \theta_0$. Therefore, the inequality $l \geqslant c$ is equivalent to $T \geqslant t$.

Let α be the given probability of Type I error. We find the integer t_α from the condition

$$\alpha'' = \sum_{m=t_\alpha+1}^{kn} C_{kn}^m\theta_0^m(1-\theta_0)^{kn-m} < \alpha \leqslant \sum_{m=t_\alpha}^{kn} C_{kn}^m\theta_0^m(1-\theta_0)^{kn-m} = \alpha'.\ (*)$$

If $\alpha = \alpha'$, the sought-for test is non-randomized and has the form $\mathscr{X}_{1\alpha}^* = \{T \geqslant t_\alpha\}$.

Since $\mathscr{L}_\theta(T) = Bi(kn,\ \theta)$ (see Problem 1.39 (3)), the probability of Type I error for this test is

$$W(\mathscr{X}_{1\alpha}^*;\ \theta_0) = \mathbf{P}_{\theta_0}(T \geqslant t_\alpha) = \alpha' = \alpha,$$

and its power is

$$W(\mathscr{X}^*_{1\alpha};\ \theta_1) = \mathbf{P}_{\theta_1}(T \geqslant t_\alpha) = \sum_{m=t_\alpha}^{kn} C^m_{kn}\theta_1^m(1-\theta_1)^{kn-m}.$$

If we have a strict inequality in $(*)$, i.e., $\alpha < \alpha'$, then the Neyman-Pearson test is randomized and defined by the critical function

$$\varphi^*_\alpha(T) = \begin{cases} 1 & \text{for } T \geqslant t_\alpha + 1, \\ \dfrac{\alpha - \alpha''}{\alpha' - \alpha''} & \text{for } T = t_\alpha, \\ 0 & \text{for } T \leqslant t_\alpha - 1. \end{cases}$$

The test power is

$$W(\varphi^*_\alpha;\ \theta_1) = \mathbf{E}_{\theta_1}\varphi^*_\alpha(T) = \mathbf{P}_{\theta_1}(T \geqslant t_\alpha + 1) + \frac{\alpha - \alpha''}{\alpha' - \alpha''}\mathbf{P}_{\theta_1}(T = t_\alpha)$$

$$= \sum_{m=t_\alpha+1}^{kn} C^m_{kn}\theta_1^m(1-\theta_1)^{kn-m}$$

$$+ (\alpha - \alpha'')\left(\frac{\theta_1}{\theta_0}\right)^{t_\alpha}\left(\frac{1-\theta_1}{1-\theta_0}\right)^{kn-t_\alpha}.$$

In this case (for $\alpha < \alpha'$) we may also use non-randomized most powerful tests $\mathscr{X}^*_{1\alpha'} = \{T \geqslant t_\alpha\}$ and $\mathscr{X}^*_{1\alpha''} = \{T \geqslant t_\alpha + 1\}$ with the significance levels $\alpha' > \alpha$ and $\alpha'' < \alpha$, respectively.

3.40. By the De Moivre-Laplace theorem we have

$$\mathscr{L}_\theta(T) \sim \mathcal{N}(kn\theta,\ kn\theta(1-\theta))$$

as $n \to \infty$. Therefore, the condition $(*)$ to define the critical boundary t_α may be replaced by an approximate condition

$$\mathbf{P}_{\theta_0}(T \geqslant t_\alpha) \simeq \Phi\left(\frac{kn\theta_0 - t_\alpha}{\sqrt{kn\theta_0(1-\theta_0)}}\right) = \alpha,$$

whence follows the sought-for critical region. We then have

$$W_n(\theta_1^{(n)}) = \mathbf{P}_{\theta_1^{(n)}}(T \geqslant t_\alpha) = \mathbf{P}_{\theta_1^{(n)}}\left(\frac{T - kn\theta_1^{(n)}}{\sqrt{kn\theta_1^{(n)}(1-\theta_1^{(n)})}}\right.$$

$$\left.\geqslant -\sqrt{\frac{kn}{\theta_1^{(n)}(1-\theta_1^{(n)})}}(\theta_1^{(n)} - \theta_0) - u_\alpha\sqrt{\frac{\theta_0(1-\theta_0)}{\theta_1^{(n)}(1-\theta_1^{(n)})}}\right)$$

$$= \Phi\left(\beta\sqrt{\frac{k}{\theta_0(1-\theta_0)}} + u_\alpha + o(1)\right) + o(1),$$

which is equivalent to the second assertion.

3.41. The test is constructed by the scheme of Problem 3.39. Here the sufficient statistic has the form $T = \sum_{i=1}^{n} X_i$ and $\mathcal{L}_\theta(T) = \Pi(n\theta)$, $l(\mathbf{X}) = \left(\frac{\theta_1}{\theta_0}\right)^T e^{n(\theta_0 - \theta_1)}$. The condition

$$\alpha'' = \sum_{m=t_\alpha+1}^{\infty} e^{-n\theta_0} \frac{(n\theta_0)^m}{m!} < \alpha \leqslant \sum_{m=t_\alpha}^{\infty} e^{-n\theta_0} \frac{(n\theta_0)^m}{m!} = \alpha' \qquad (*)$$

allows us to find the critical boundary t_α for the given probability α of Type I error. For $\alpha = \alpha'$ the test has the form $\mathscr{X}^*_{1\alpha} = \{T \geqslant t_\alpha\}$, and for $\alpha < \alpha'$ it is randomized and its critical function $\varphi^*_\alpha(T)$ has the form given in Problem 3.39 (taking into account the notations from (*)). In any case the power is computed by the formula

$$W(\theta_1) = \sum_{m=t_\alpha+1}^{\infty} e^{-n\theta_1} \frac{(n\theta_1)^m}{m!} + (\alpha - \alpha'')\left(\frac{\theta_1}{\theta_0}\right)^{t_\alpha} e^{-n(\theta_1 - \theta_0)}.$$

If $n \to \infty$, then $\mathcal{L}_\theta(T) \sim \mathcal{N}(n\theta, n\theta)$ and, by reasoning as in Problem 3.40, we find that the test asymptotic form is $\{T \geqslant n\theta_0 - u_\alpha\sqrt{n\theta_0}\}$, and its power under the close alternative $\theta_1 = \theta_1^{(n)} = \theta_0 + \beta/\sqrt{n}$, $\beta > 0$, satisfies the limiting relation

$$\lim_{n\to\infty} W_n(\theta_1^{(n)}) = \Phi\left(\frac{\beta}{\sqrt{\theta_0}} + u_\alpha\right).$$

3.42. The observable random variable X has a geometric distribution $\overline{Bi}(1, \theta)$, and the likelihood ratio statistic is $l(X) = \left(\frac{\theta_1}{\theta_0}\right)^X \frac{1 - \theta_1}{1 - \theta_0}$. Consequently, the inequality $l \geqslant c$ is equivalent to $X \geqslant t$, and the critical boundary $t = t_\alpha$ is found from

$$\alpha = \theta_0^s = \mathbf{P}_{\theta_0}(X \geqslant t_\alpha) = \sum_{m=t_\alpha}^{\infty} \theta_0^m(1 - \theta_0) = \theta_0^{t_\alpha},$$

whence $t_\alpha = s$.

In our case the Neyman-Pearson test has the form $\mathscr{X}^*_{1\alpha} = \{X \geqslant s\}$, and its power is

$$1 - \beta = W(\theta_1) = \mathbf{P}_{\theta_1}(X \geqslant s) = \sum_{m=s}^{\infty} \theta_1^m(1 - \theta_1) = \theta_1^s.$$

3.43. Here the likelihood ratio statistic is

$$l(\mathbf{X}) = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left\{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) T\right\}, \quad T = \sum_{i=1}^{n} X_i,$$

and $\mathscr{L}(2T/\theta) = \chi^2(2n)$ (see the hint). If $\theta_0 < \theta_1$, the inequality $l \geqslant c$ is equivalent to $T \geqslant t$, and the condition for finding the critical boundary $t = t_\alpha$ at the significance level α is

$$\alpha = \mathbf{P}_{\theta_0}(T \geqslant t_\alpha) = \mathbf{P}_{\theta_0}(2T/\theta_0 \geqslant 2t_\alpha/\theta_0) = \mathbf{P}(\chi^2_{2n} \geqslant 2t_\alpha/\theta_0).$$

Then $2t_\alpha/\theta_0 = \chi^2_{1-\alpha,2n}$, and the optimal test has the form $\mathscr{X}^*_{1\alpha} = \left\{T \geqslant \dfrac{\theta_0}{2}\chi^2_{1-\alpha,2n}\right\}$. Its power is

$$W(\theta_1) = \mathbf{P}_{\theta_1}\left(T \geqslant \frac{\theta_0}{2}\chi^2_{1-\alpha,2n}\right) = \mathbf{P}_{\theta_1}(2T/\theta_1 \geqslant \theta_0\chi^2_{1-\alpha,2n}/\theta_1)$$

$$= 1 - F_{2n}\left(\frac{\theta_0}{\theta_1}\chi^2_{1-\alpha,2n}\right),$$

where $F_{2n}(t)$ is the distribution function of the law $\chi^2(2n)$. Similarly, for $\theta_0 > \theta_1$ the optimal test has the form $\mathscr{X}^*_{1\alpha} = \left\{T \leqslant \dfrac{\theta_0}{2}\chi^2_{\alpha,2n}\right\}$, and its power is

$$W(\theta_1) = F_{2n}\left(\frac{\theta_0}{\theta_1}\chi^2_{\alpha,2n}\right).$$

3.44. In our case the set of critical values for the observation is found from

$$l(x) = \frac{1 + x^2}{1 + (x-1)^2} \geqslant c,$$

which is equivalent to the condition $(c-1)x^2 - 2cx + 2c - 1 \leqslant 0$. If $c = 1$, the inequality holds for $x \geqslant 1/2$. This means that the $\mathscr{X}^*_{1\alpha} = \{X \geqslant 1/2\}$-test has the significance level

$$\alpha = \mathbf{P}_0\left(X \geqslant \frac{1}{2}\right) = \frac{1}{\pi}\int_{1/2}^{\infty} \frac{dx}{1 + x^2} = \frac{1}{2} - \frac{1}{\pi}\arctan\frac{1}{2},$$

and its power is

$$\mathbf{P}_1\left(X \geqslant \frac{1}{2}\right) = \frac{1}{\pi}\int_{1/2}^{\infty} \frac{dx}{1 + (x-1)^2} = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{1}{2}.$$

Putting $c = 2$, we obtain the inequality $(x - 2)^2 \leqslant 1$ or $1 \leqslant x \leqslant 3$. This means that the significance level of the $\mathscr{X}^*_{1\alpha} = \{1 \leqslant X \leqslant 3\}$-test is

$$\alpha = \mathbf{P}_0(1 \leqslant X \leqslant 3) = \frac{1}{\pi}\int_1^3 \frac{dx}{1 + x^2} = \frac{1}{\pi}(\arctan 3 - \arctan 1),$$

and its power is

$$\mathbf{P}_1(1 \leqslant X \leqslant 3) = \frac{1}{\pi}\int_1^3 \frac{dx}{1 + (x - 1)^2} = \frac{1}{\pi}\arctan 2.$$

3.45. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from the distribution $\mathscr{L}(\xi)$. If we have at least one $|X_i| > a$, then this event is impossible under the hypothesis H_0 and it must be rejected. In other cases, i.e., for $T_n^{(1)} \equiv \max\limits_{1 \leqslant i \leqslant n} |X_i| \leqslant a$, we make our decision by investigating the likelihood ratio statistic

$$l(\mathbf{X}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-X_i^2/2\sigma^2} \Big/ \frac{1}{(2a)^n} = \left(\sqrt{\frac{2}{\pi}}\frac{a}{\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}T_n^{(2)}\right\},$$

$$T_n^{(2)} = \sum_{i=1}^{n} X_i^2.$$

Here the inequality $l \geqslant c$ is equivalent to $T_n^{(2)} \leqslant t$, and the sought-for test has the form

$$\begin{aligned}\mathscr{X}^*_{1\alpha} &= \{T_n^{(1)} > a\} \cup \{T_n^{(1)} \leqslant a, \quad T_n^{(2)} \leqslant t_\alpha\} \\ &= \{T_n^{(1)} > a \quad \text{or} \quad T_n^{(2)} \leqslant t_\alpha\},\end{aligned}$$

where the boundary t_α at the significance level α is found from the condition

$$\alpha = \mathbf{P}(\mathbf{X} \in \mathscr{X}^*_{1\alpha} | H_0) = \mathbf{P}(T_n^{(2)} \leqslant t_\alpha | H_0)$$

$$= \int\cdots\int_{x_1^2 + \ldots + x_n^2 \leqslant t_\alpha} \frac{dx_1 \ldots dx_n}{(2a)^n} = \frac{(\pi t_\alpha)^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)(2a)^n}$$

(because the event $\{T_n^{(1)} \leqslant a\}$ is certain under the hypothesis H_0). We find the estimate

$$\begin{aligned}W(\mathscr{X}^*_{1\alpha}; H_1) \geqslant \mathbf{P}(T_n^{(1)} > a | H_1) &= 1 - \mathbf{P}^n(|X_1| \leqslant a | H_1) \\ &= 1 - \left[1 - 2\Phi\left(-\frac{a}{\sigma}\right)\right]^n\end{aligned}$$

for the power of the test. Consequently, the probability β of Type II error satisfies the inequality

$$\beta \leqslant \left[1 - 2\Phi\left(-\frac{a}{\sigma}\right)\right]^n$$

for any α. If $n \to \infty$, by the Central Limit Theorem we have

$$\mathscr{L}(T_n^{(2)} | H_0) \sim \mathscr{N}(n\mu, nb^2),$$

where

$$\mu = \mathbf{E}(X_1^2 | H_0) = \frac{1}{2a} \int_{-a}^{a} x^2\, dx = \frac{a^2}{3},$$

$$b^2 = \mathbf{D}(X_1^2 | H_0) = \frac{1}{2a} \int_{-a}^{a} x^4\, dx - \mu^2 = \frac{4}{45} a^4.$$

Consequently, we may use an approximate equation

$$\alpha \simeq \Phi\left(\left(t_\alpha - \frac{na^2}{3}\right) \Big/ \sqrt{\frac{4na^4}{45}}\right)$$

to find t_α, whence

$$t_\alpha = \frac{na^2}{3} + u_\alpha \frac{2a^2}{3} \sqrt{\frac{n}{5}}.$$

3.46. We use T_n to denote the number of positive outcomes in n trials. Then $\mathscr{L}_p(T_n) = Bi(n, p)$, and the event $\{T_n > 0\}$ is impossible under the hypothesis H_0. It follows that the test should be given as $\mathscr{X}_1 = \{T_n > 0\}$, i.e., we accept the hypothesis H_0 for $T_n = 0$ and accept the hypothesis H_1 in all the other cases. Then the probabilities of errors will be

$$\alpha = \mathbf{P}(T_n > 0 | H_0) = 0, \quad \beta = \mathbf{P}(T_n = 0 | H_1) = 0.99^n.$$

We find n from the condition $0.99^n \leqslant 0.01$, whence $n \geqslant 454$.

3.47. Here the likelihood ratio statistic is reduced to the form

$$l(\mathbf{X}) = \exp\left\{\frac{n}{\sigma^2}(\theta_1 - \theta_0)\overline{X} - \frac{n}{2\sigma^2}(\theta_1^2 - \theta_0^2)\right\}.$$

If $\theta_1 > \theta_0$, then the critical set $\mathscr{X}^*_{1\alpha}$ has the form

$$\mathscr{X}^*_{1\alpha} = \left\{\bar{x} \geqslant \theta_0 - \frac{\sigma}{\sqrt{n}} u_\alpha\right\},$$

and its power is

$$W(\theta_1) = \mathbf{P}_{\theta_1}\left(\overline{X} \geqslant \theta_0 - \frac{\sigma}{\sqrt{n}} u_\alpha\right) = \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta_1 - \theta_0) + u_\alpha\right) > \alpha.$$

If $\theta_0 > \theta_1$, then we have

$$\mathscr{S}^*_{1\alpha} = \left\{\bar{x} \leqslant \theta_0 + \frac{\sigma}{\sqrt{n}} u_\alpha\right\}; \quad W(\theta_1) = \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta_0 - \theta_1) + u_\alpha\right).$$

3.48. We have two equations

$$\Phi(u_\alpha) = \alpha, \quad \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta_1 - \theta_0) + u_\alpha\right) = 1 - \beta$$

to find n, whence follows that n^* is the minimal integer which is not smaller than $\sigma^2(u_\alpha + u_\beta)^2/(\theta_1 - \theta_0)^2$.

3.49. Since $\mathscr{L}(T|H_0) = \mathscr{N}(0, 1)$, $\mathscr{L}(T|H_1) = \mathscr{N}(\Delta/\sigma, 1)$, we are dealing with two simple hypotheses about the mean of the normal distribution with variance equal to unity, on the basis of one observation on the random variable T.

The solution to Problem 3.47 gives the sought-for $\mathscr{S}^*_{1\alpha} = \{T > -u_\alpha\}$-test with $\beta = \Phi(-u_\alpha - \Delta/\sigma)$. Taking into account that $u_\alpha = \Phi^{-1}(\alpha)$, we find m from the equation

$$-u_\alpha - \Delta/\sigma = \Phi^{-1}(\beta) = u_\beta \quad \text{or} \quad \sigma^2 = \Delta^2(u_\alpha + u_\beta)^{-2}.$$

Finally, solving the latter equality with respect to m, we conclude that m^* is the minimal integer no smaller than

$$(u_\alpha + u_\beta)^2 \Big/ \left[\frac{\Delta^2}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma_2^2}(u_\alpha + u_\beta)^2\right].$$

3.50. Let H_i: $\theta = \theta_i$, $i = 0, 1$, and $\theta_0 > \theta_1$. If $\mathbf{X} = (X_1, \ldots, X_n)$ is the required sample, then the likelihood ratio statistic is

$$l(\mathbf{X}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\theta_1} e^{-\frac{1}{2\theta_1^2}(X_i - \mu)^2} \Big/ \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\theta_0} e^{-\frac{1}{2\theta_0^2}(X_i - \mu)^2}$$

$$= \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left\{-\frac{\theta_0^2 - \theta_1^2}{2\theta_0^2\theta_1^2} T\right\}, \quad T = \sum_{i=1}^{n}(X_i - \mu)^2,$$

where $\mathscr{L}_\theta(T/\theta^2) = \chi^2(n)$. The inequality $l \geqslant c$ is equivalent to $T \leqslant t$. Therefore, the critical boundary $t = t_\alpha$ at the significance level α is found from the

condition

$$\alpha = \mathbf{P}_{\theta_0}(T \leqslant t_\alpha) = \mathbf{P}_{\theta_0}(T/\theta_0^2 \leqslant t_\alpha/\theta_0^2) = F_n(t_\alpha/\theta_0^2),$$

where $F_n(t)$ is the distribution function of the law $\chi^2(n)$. Then $t_\alpha/\theta_0^2 = \chi^2_{\alpha,n}$, and the sought-for test has the form

$$\mathscr{X}^*_{1\alpha} = \{T \leqslant \theta_0^2\chi^2_{\alpha,n}\}.$$

Its power is

$$W(\theta_1) = \mathbf{P}_{\theta_1}(T \leqslant \theta_0^2\chi^2_{\alpha,n}) = \mathbf{P}_{\theta_1}\left(T/\theta_1^2 \leqslant \left(\frac{\theta_0}{\theta_1}\right)^2\chi^2_{\alpha,n}\right) = F_n\left(\left(\frac{\theta_0}{\theta_1}\right)^2\chi^2_{\alpha,n}\right).$$

For $\theta_0 < \theta_1$ we find in a similar way that

$$\mathscr{X}^*_{1\alpha} = \{T \geqslant \theta_0^2\chi^2_{1-\alpha,n}\}, \quad W(\theta_1) = 1 - F_n\left(\left(\frac{\theta_0}{\theta_1}\right)^2\chi^2_{1-\alpha,n}\right).$$

3.51. (1) The assertion follows from

$$\alpha(c) = \int_{\mathscr{X}_1^*(c)} f_0(x)\,dx \leqslant \frac{1}{c}\int_{\mathscr{X}_1^*(c)} f_1(x)\,dx = \frac{1}{c}(1 - \beta(c)),$$

$$\beta(c) = \int_{\mathscr{X}_0^*(c)} f_1(x)\,dx \leqslant c\int_{\mathscr{X}_0^*(c)} f_0(x)\,dx = c(1 - \alpha(c)),$$

$$\mathscr{X}_0^*(c) = \overline{\mathscr{X}_1^*}(c).$$

(2) If $c > 1$, then we use the first of the above relations to find

$$\alpha(c) + \beta(c) \leqslant \frac{1}{c}(1 - \beta(c)) + \beta(c) < 1.$$

If $c \leqslant 1$, then we use the second relation to find

$$\alpha(c) + \beta(c) \leqslant \alpha(c) + c(1 - \alpha(c)) \leqslant 1.$$

(3) If $c > 1$, then we use the relation

$$\alpha(c) + \beta(c) = \int_{\mathscr{X}_1^*(c)} f_0(x)\,dx + \int_{\mathscr{X}_0^*(c)} f_1(x)\,dx$$

$$= 1 - \int_{\mathscr{X}_1^*(c)} (f_1(x) - f_0(x))\,dx.$$

It is clear that $\mathscr{S}_1(c) \subseteq \mathscr{S}_1(1)$, and we have $f_1(x) - f_0(x) \geqslant 0$ on the set $\mathscr{S}_1(1)$. Consequently,

$$\int_{\mathscr{S}_1(c)} (f_1(x) - f_0(x))\,dx \leqslant \int_{\mathscr{S}_1(1)} (f_1(x) - f_0(x))\,dx,$$

whence $\alpha(c) + \beta(c) \geqslant \alpha(1) + \beta(1)$.

If $c < 1$, then we proceed from the relation

$$\alpha(c) + \beta(c) = 1 - \int_{\mathscr{S}_0(c)} (f_0(x) - f_1(x))\,dx.$$

(4) If the hypothesis H_0 is true, then by the law of large numbers we have

$$T_n(X) \xrightarrow{\mathbf{P}} \mathbf{E}\left(\ln \frac{f_1(X_1)}{f_0(X_1)} \middle| H_0\right) = \delta$$

as $n \to \infty$. Therefore, for $\delta < 0$ we have

$$\alpha_n = \mathbf{P}(T_n(X) \geqslant 0 | H_0) \leqslant \mathbf{P}(|T_n(X) - \delta| \geqslant |\delta| \,|\, H_0) \to 0$$

as $n \to \infty$. By symmetry (if we change H_0 for H_1 and vice versa), β_n tends to zero.

3.52. In this case

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{r/2}\sqrt{|\mathbf{A}|}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}^{(i)})'\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}^{(i)})\right\}, \quad i = 0, 1,$$

and the region $\mathscr{S}_1(c) = \{\mathbf{x}: f_1(\mathbf{x})/f_0(\mathbf{x}) \geqslant c\}$ has the form

$$\mathscr{S}_1(c) = \left\{\mathbf{x}: \mathbf{a}'\mathbf{x} - \frac{1}{2}\mathbf{a}'(\boldsymbol{\mu}^{(0)} + \boldsymbol{\mu}^{(1)}) \leqslant c_1 = -\ln c\right\},$$

$$\mathbf{a} = \mathbf{A}^{-1}(\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu}^{(1)}).$$

For the random variable $Y = \mathbf{a}'\xi - \frac{1}{2}\mathbf{a}'(\boldsymbol{\mu}^{(0)} + \boldsymbol{\mu}^{(1)})$ we have

$$\mathscr{L}(Y|H_i) = \mathscr{N}\left((-1)^i \frac{\varrho}{2}, \varrho\right),$$

where $\varrho = (\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu}^{(1)})'\mathbf{A}^{-1}(\boldsymbol{\mu}^{(0)} - \boldsymbol{\mu}^{(1)})$ is the Mahalanobis distance between the distributions $\mathscr{N}(\boldsymbol{\mu}^{(0)}, \mathbf{A})$ and $\mathscr{N}(\boldsymbol{\mu}^{(1)}, \mathbf{A})$. We now find the probabilities of Type I and Type II errors for the $\mathscr{S}_1(c)$-test. They are

$$\alpha(c) = \mathbf{P}(Y \leqslant c_1 | H_0) = \Phi\left(\frac{c_1 - \varrho/2}{\sqrt{\varrho}}\right),$$

$$\beta(c) = \mathbf{P}(Y > c_1 | H_1) = \Phi\left(-\frac{c_1 + \varrho/2}{\sqrt{\varrho}}\right).$$

If the probability α of Type I error is given, then the respective Neyman-Pearson test is defined by the critical region $\mathscr{X}_1(c)$ for $c_1 = \varrho/2 + \sqrt{\varrho}u_\alpha$, $u_\alpha = \Phi^{-1}(\alpha)$, the probability of Type II error being equal to $\beta = \Phi(-\sqrt{\varrho} - u_\alpha)$. The $\mathscr{X}_1(1)$-test minimizes the sum of the probabilities of errors (see Problem 3.51) and this sum is $2\Phi(-\sqrt{\varrho}/2)$.

3.53. The solution of Problem 3.39 implies that the model's likelihood ratio is monotone, and we may use the Neyman-Pearson test from Problem 3.39 as a u.m.p. test for the given problem.

3.55. Since $\mathscr{L}_\theta(T_r) = \overline{Bi}(r, \theta)$ is an exponential distribution with a monotone increasing function $A(\theta) = \ln \theta$, the u.m.p. test exists and is defined by the critical region of the form $\{T_r \geqslant t\}$ (see Sec. 3.5). By the Central Limit Theorem we have $\mathscr{L}_\theta(T_r) \sim \mathscr{N}\left(\frac{r\theta}{1-\theta}, \frac{r\theta}{(1-\theta)^2}\right)$ as $r \to \infty$, and we may use the relations

$$\alpha = \mathbf{P}_{\theta_0}(T_r \geqslant t_\alpha) \simeq \Phi\left(\left(\frac{r\theta_0}{1-\theta_0} - t_\alpha\right) \Big/ \sqrt{\frac{r\theta_0}{(1-\theta_0)^2}}\right)$$

to calculate the critical boundary $t = t_\alpha$ for large r. We then find the required formula for t_α.

3.57. We have

$$\mathbf{P}_\theta(T = x) = f(x; \theta) = C_\theta^x C_{N-\theta}^{n-x} / C_N^n, \quad x = 0, 1, \ldots, \theta,$$

and therefore the function

$$l(x) = \frac{f(x; \theta + 1)}{f(x; \theta)} = \frac{\theta + 1}{N - \theta} \frac{N - \theta - n + x}{\theta + 1 - x}$$

is monotone increasing in x. Consequently, the u.m.p. exists and has the form $\{T \geqslant t\}$, i.e., the hypothesis H_0 is rejected when T is too large.

3.60. Consider the class

$$\mathscr{X}_{1\alpha} = \left\{\frac{\sqrt{n}}{\sigma}(\bar{x} - \theta_0) \leqslant u_{\alpha_1}\right\} \cup \left\{\frac{\sqrt{n}}{\sigma}(\bar{x} - \theta_0) \geqslant -u_{\alpha_2}\right\},$$

where $\alpha_1 + \alpha_2 = \alpha$, and the power function is

$$W(\mathscr{X}_{1\alpha}; \theta) = \Phi(\sqrt{n}\Delta/\sigma + u_{\alpha_2}) + \Phi(-\sqrt{n}\Delta/\sigma + u_{\alpha_1}), \quad \Delta = \theta - \theta_0.$$

Clearly, the power is minimal for $\Delta = \Delta_0 = \sigma(u_{\alpha_1} - u_{\alpha_2})/(2\sqrt{n})$. Since

$W(\mathscr{S}_{1\alpha};\ \theta_0) = \alpha$ (for $\Delta = 0$), the test is unbiased only for $\Delta_0 = 0$, i.e., for $\alpha_1 = \alpha_2 = \alpha/2$. We then find the sought-for test

$$\tilde{\mathscr{S}}_{1\alpha} = \left\{\frac{\sqrt{n}}{\sigma}|\bar{x} - \theta_0| \geqslant -u_{\alpha/2}\right\}.$$

This is the u.m.p. test among all the unbiased tests.

3.61. The likelihood function is

$$L(\mathbf{x};\ \theta) = \frac{1}{(\sqrt{2\pi}\,\theta)^n}\,e^{-T(\mathbf{x})/2\theta^2},\quad T(\mathbf{x}) = \sum_{i=1}^{n}(x_i - \mu)^2,$$

$$L_1(\mathbf{x};\ \theta) = \frac{\partial L(\mathbf{x};\ \theta)}{\partial\theta} = \left(\frac{T(\mathbf{x})}{\theta^3} - \frac{n}{\theta}\right)L(\mathbf{x};\ \theta).$$

Therefore, the inequality $L(\mathbf{x};\ \theta_1) \geqslant cL(\mathbf{x};\ \theta_0) + c_1L_1(\mathbf{x};\ \theta_0)$, which defines the best critical region, takes the form (see the solution to Problem 3.50)

$$e^{aT} \geqslant c' + c_1'T \quad \text{or} \quad \{T \leqslant t_1\} \cup \{T \geqslant t_2\}.$$

Thus, the sought-for test is

$$\tilde{\mathscr{S}}_1 = \{T \leqslant t_1\} \cup \{T \geqslant t_2\},$$

where the boundaries $t_1 < t_2$ are defined at the given significance level α by the conditions $W(\theta_0) = \alpha$, $W'(\theta_0) = 0$, with the power function $W(\theta)$. We know (see the solution to Problem 3.50) that

$$W(\theta) = \mathbf{P}_\theta(\mathscr{X} \in \tilde{\mathscr{S}}_1) = \mathbf{P}_\theta(T \leqslant t_1) + \mathbf{P}_\theta(T \geqslant t_2)$$
$$= F_n(t_1/\theta^2) + 1 - F_n(t_2/\theta^2),$$

whence we have two equations

$$F_n(t_1/\theta_0^2) + 1 - F_n(t_2/\theta_0^2) = \alpha,\quad t_1k_n(t_1/\theta_0^2) = t_2k_n(t_2/\theta_0^2),$$

where $k_n(t) = F_n'(t)$. Putting $t_1 = \theta_0^2\chi^2_{\alpha_1,n}$, $t_2 = \theta_0^2\chi^2_{1-\alpha_2,n}$, we find that the conditions

$$\chi^2_{\alpha_1,n}k_n(\chi^2_{\alpha_1,n}) = \chi^2_{1-\alpha_2,n}k_n(\chi^2_{1-\alpha_2,n}),\quad \alpha_1 + \alpha_2 = \alpha,$$

should be met. These conditions uniquely define $\chi^2_{\alpha_1,n}$ and $\chi^2_{1-\alpha_2,n}$, and hence t_1 and t_2. The sought-for test takes the form

$$\tilde{\mathscr{S}}_{1\alpha} = \{T \leqslant \theta_0^2\chi^2_{\alpha_1,n}\} \cup \{T \geqslant \theta_0^2\chi^2_{1-\alpha_2,n}\}$$

and is the union of two one-sided u.m.p. tests from Problem 3.59. The values of $(\chi^2_{\alpha_1,n},\ \chi^2_{1-\alpha_2,n})$ for $\alpha = 0.05$ and $n = 2, 5, 10, 20$ can be found in [7, p. 108].

3.62. The algorithm of solution is here as in Problem 3.61. Using the result and notations from Problem 3.43, we find that the test is of the form

$$\tilde{\mathscr{S}}_1 = \{T \leqslant t_1\} \cup \{T \geqslant t_2\},$$

and its power function is

$$W(\theta) = F_{2n}(2t_1/\theta) + 1 - F_{2n}(2t_2/\theta).$$

We have two equations $W(\theta_0) = \alpha$ and $W'(\theta_0) = 0$ to define the boundaries t_1 and t_2. For $t_1 = \frac{\theta_0}{2}\chi^2_{\alpha_1, 2n}$, $t_2 = \frac{\theta_0}{2}\chi^2_{1-\alpha_2, 2n}$ the equations take the form

$$\chi^2_{\alpha_1, 2n} k_{2n}(\chi^2_{\alpha_1, 2n}) = \chi^2_{1-\alpha_2, 2n} k_{2n}(\chi^2_{1-\alpha_2, 2n}), \quad \alpha_1 + \alpha_2 = \alpha.$$

We find $\chi^2_{\alpha_1, 2n}$ and $\chi^2_{1-\alpha_2, 2n}$ and conclude that at the significance level α the sought-for test has the form

$$\tilde{\mathscr{X}}_{1\alpha} = \left\{ \sum_{i=1}^{n} x_i \leqslant \frac{\theta_0}{2}\chi^2_{\alpha_1, 2n} \quad \text{or} \quad \sum_{i=1}^{n} x_i \geqslant \frac{\theta_0}{2}\chi^2_{1-\alpha_2, 2n} \right\}$$

and is a union of two respective one-sided u.m.p. tests for the alternatives H_1^-: $\theta < \theta_0$ and H_1^+: $\theta > \theta_0$ which follows from Problem 3.43 and the properties of the exponential model.

3.63. In our case (see the solution to Problem 3.39) the likelihood function is

$$L(\mathbf{x};\, \theta) = \prod_{i=1}^{n} C_k^{x_i} \theta^{x_i} (1 - \theta)^{k - x_i}.$$

Therefore,

$$U(\mathbf{x};\, \theta) = \frac{\partial \ln L(\mathbf{x};\, \theta)}{\partial \theta} = \frac{T(\mathbf{x}) - kn\theta}{\theta(1 - \theta)}, \quad T(\mathbf{x}) = \sum_{i=1}^{n} x_i.$$

Besides (see Problem 2.43), we have $i(\theta) = k/[\theta(1 - \theta)]$, and the sought-for test (see the hint) has the form

$$\mathscr{X}_{1\alpha} = \{|T - kn\theta_0|/\sqrt{kn\theta_0(1 - \theta_0)} \geqslant -u_{\alpha/2}\}.$$

The test's power for the given alternatives is calculated as in Problem 3.40.

3.64. We have

$$L(\mathbf{x};\, \theta) = e^{-n\theta}\theta^{T(\mathbf{x})}/(x_1! \ldots x_n!), \; T(\mathbf{x}) = \sum_{i=1}^{n} x_i,$$

and hence $U(\mathbf{x};\, \theta) = (T(\mathbf{x}) - n\theta)/\theta$, $i(\theta) = 1/\theta$ (see Problem 2.43). The sought-for test has the form

$$\mathscr{X}_{1\alpha} = \{|T - n\theta_0|/\sqrt{n\theta_0} \geqslant -u_{\alpha/2}\}.$$

Its power is calculated as in Problem 3.41.

3.65. We use the notations from Problem 2.119 and find that the critical regions $\mathscr{S}_{1\alpha}$ for our problems have the forms

(1) $\mathscr{S}_{1\alpha} = \{\mathbf{x}: \bar{x} \geqslant \theta_{10} + t_{1-\alpha, n-1} S(\mathbf{x})/\sqrt{n-1}\}$;

(2) $\mathscr{S}_{1\alpha} = \{\mathbf{x}: \bar{x} \leqslant \theta_{10} - t_{1-\alpha, n-1} S(\mathbf{x})/\sqrt{n-1}\}$ (compare with the results of Problems 3.47 and 3.58 for the case when the variance is known);

(3) $\mathscr{S}_{1\alpha} = \left\{\mathbf{x}: \sqrt{n-1}\,\dfrac{|\bar{x} - \theta_{10}|}{S(\mathbf{x})} \geqslant t_{1-\alpha/2, n-1}\right\}$ (compare with the results of Problems 3.60 and 3.73);

(4) $\mathscr{S}_{1\alpha} = \{\mathbf{x}: nS^2(\mathbf{x}) \geqslant \theta_{20}^2 \chi^2_{1-\alpha, n-1}\}$;

(5) $\mathscr{S}_{1\alpha} = \{\mathbf{x}: nS^2(\mathbf{x}) \leqslant \theta_{20}^2 \chi^2_{\alpha, n-1}\}$ (compare with the results of Problems 3.50 and 3.59);

(6) $\mathscr{S}_{1\alpha} = \{\mathbf{x}: nS^2(\mathbf{x}) \leqslant \theta_{20}^2 \chi^2_{\alpha/2, n-1}$ or $nS^2(\mathbf{x}) \geqslant \theta_{20}^2 \chi^2_{1-\alpha/2, n-1}\}$ (compare with the results of Problems 3.61 and 3.74).

Indeed, let us investigate the first problem (all the other problems are solved in the same way). We use the lower γ-confidence interval $\mathscr{I}_\gamma(\mathbf{X}) = \{\theta_1: \overline{X} - t_{\gamma, n-1} S(\mathbf{X})/\sqrt{n-1} < \theta_1 < \infty\}$ for θ_1 and find that the acceptance region for the hypothesis H_0 with the significance level $\alpha = 1 - \gamma$ has the form

$$\mathscr{S}_{0\alpha} = \{\mathbf{x}: \bar{x} - t_{\gamma, n-1} S(\mathbf{x})/\sqrt{n-1} < \theta_{10}\}.$$

But $\mathscr{S}_{1\alpha} = \overline{\mathscr{X}_{0\alpha}}$ is as given in the problem.

3.68. Using the γ-confidence interval for τ constructed in Problem 2.127, we find that the acceptance region for the hypothesis H_0 is

$$\mathscr{S}_{0\alpha} = \left\{(\mathbf{x}, \mathbf{y}): \frac{\bar{y}}{\bar{x}} F_{(1-\gamma)/2, 2n, 2m} < 1 < \frac{\bar{y}}{\bar{x}} F_{(1+\gamma)/2, 2n, 2m}\right\}, \quad \alpha = 1 - \gamma.$$

The required test is defined by the critical region

$$\mathscr{S}_{1\alpha} = \left\{(\mathbf{x}, \mathbf{y}): \frac{\bar{x}}{\bar{y}} \leqslant F_{\alpha/2, 2n, 2m} \quad \text{or} \quad \frac{\bar{x}}{\bar{y}} \geqslant F_{1-\alpha/2, 2n, 2m}\right\},$$

and its power is

$$W(\theta) = \mathbf{P}_\theta\left(\tau \frac{\overline{X}}{\overline{Y}} \leqslant \tau F_{\alpha/2, 2n, 2m}\right) + \mathbf{P}_\theta\left(\tau \frac{\overline{X}}{\overline{Y}} \geqslant \tau F_{1-\alpha/2, 2n, 2m}\right)$$

$$= F(\tau F_{\alpha/2, 2n, 2m};\ 2n,\ 2m) + 1 - F(\tau F_{1-\alpha/2, 2n, 2m};\ 2n,\ 2m), \quad \tau = \frac{\theta_2}{\theta_1}.$$

3.69. We invert the γ-confidence interval constructed in Problem 2.128 and find the critical set for the hypothesis H_0, viz.,

$$\mathscr{S}_{1\alpha} = \{\mathbf{x}: x_{(1)} \leqslant \theta_0 \quad \text{or} \quad x_{(1)} \geqslant \theta_0 - (\ln \alpha)/n\}.$$

Since

$$\mathbf{P}_\theta(X_{(1)} \geqslant t) = \begin{cases} 1 & \text{for} \quad t < \theta, \\ e^{-n(t-\theta)} & \text{for} \quad t \geqslant \theta \end{cases}$$

(see the solution to Problem 2.128), the power function of the test is

$$W(\theta) = \mathbf{P}_\theta(X_{(1)} \leqslant \theta_0) + \mathbf{P}_\theta(X_{(1)} \geqslant \theta_0 - (\ln \alpha)/n)$$

$$= \begin{cases} 1, \ \theta \geqslant \theta_0 - (\ln \alpha)/n, \\ \alpha e^{n(\theta - \theta_0)}, \ \theta_0 < \theta < \theta_0 - (\ln \alpha)/n, \\ 1 - (1 - \alpha)e^{n(\theta - \theta_0)}, \ \theta \leqslant \theta_0, \end{cases}$$

whence follows that $W(\theta) \geqslant \alpha$ for all θ.

3.70. The sought-for test has the form

$$\mathscr{S}_{1\alpha} = \{\mathbf{x}: x_{(n)} \leqslant \theta_0 \alpha^{1/n} \quad \text{or} \quad x_{(n)} \geqslant \theta_0\},$$

and its power function is

$$W(\theta) = \mathbf{P}_\theta(X_{(n)} \leqslant \theta_0 \alpha^{1/n}) + \mathbf{P}_\theta(X_{(n)} \geqslant \theta_0)$$

$$= \min\left(\alpha\left(\frac{\theta_0}{\theta}\right)^n, 1\right) + 1 - \min\left(\left(\frac{\theta_0}{\theta}\right)^n, 1\right)$$

$$= \begin{cases} 1, \ \theta < \theta_0 \alpha^{1/n}, \\ \alpha\left(\dfrac{\theta_0}{\theta}\right)^n, \ \theta_0 \alpha^{1/n} \leqslant \theta \leqslant \theta_0, \\ 1 - (1 - \alpha)\left(\dfrac{\theta_0}{\theta}\right)^n, \ \theta > \theta_0, \end{cases}$$

whence it follows that $W(\theta) \geqslant \alpha$ for all θ.

3.71. The assertion of the choice of the boundaries in an unbiased test follows from the fact that the distributions of the respective statistics coincide.

3.72. Inverting the confidence region constructed in Problem 2.132, we find the critical region

$$\mathscr{S}_{1\alpha} = \{\mathbf{x}: n(\bar{x}_1 - \theta_{10}, \bar{x}_2 - \theta_{20})' \mathbf{\Sigma}^{-1}(\bar{x}_1 - \theta_{10}, \bar{x}_2 - \theta_{20}) > \chi^2_{1-\alpha, 2}\}.$$

3.73. Here $\Theta = \{\boldsymbol{\theta} = (\theta_1, \theta_2): -\infty < \theta_1 < \infty, \theta_2 > 0\}$ and (see the solution to Problem 2.86)

$$\sup_{\Theta} L(\mathbf{x}; \boldsymbol{\theta}) = L(\mathbf{x}; (\bar{x}, s)) = (2\pi e s^2)^{-n/2}, \ s^2 = S^2(\mathbf{x}).$$

We have $\Theta_0 = \{\boldsymbol{\theta} = (\theta_1, \theta_2): \theta_1 = \theta_{10}, \theta_2 > 0\}$ and

$$\sup_{\Theta_0} L(\mathbf{x}; \boldsymbol{\theta}) = L(\mathbf{x}; (\theta_{10}, s_0)) = (2\pi e s_0^2)^{-n/2},$$

where $s_0^2 = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \theta_{10})^2$ is a m.l.e. for θ_2^2 under the hypothesis H_0. Since $s_0^2 = s^2 + (\bar{x} - \theta_{10})^2$, we have

$$\lambda_n = \lambda_n(\mathbf{x}) = (s_0^2/s^2)^{-n/2} = (1 + t^2/(n - 1))^{-n/2},$$

where $t = t(\mathbf{x}) = \sqrt{n-1}(\overline{x} - \theta_{10})/s$. Therefore, the inequality $\lambda_n \leqslant c$ is equivalent to $|t| \geqslant c'$. In this case the likelihood ratio test has the form

$$\mathscr{X}_{1\alpha} = \{\mathbf{x}: \sqrt{n-1}|\overline{x} - \theta_{10}|/s \geqslant c'\}.$$

Since the test statistic $t(\mathbf{X})$ has Student's distribution $S(n-1)$ under the hypothesis H_0, the boundary will be $c' = t_{1-\alpha/2, n-1}$ (compare with Problem 3.65 (3)).

The $S(n-1)$-distribution is approximated by $\mathcal{N}(0, 1)$ for large n (see Problem 1.47). We may then take the critical boundary c' to be approximately $-u_{\alpha/2}$. Note also that $u^2_{\alpha/2} = \chi^2_{1-\alpha,1}$. The information matrix $\mathbf{I}(\theta)$ for the model $\mathcal{N}(\theta_1, \theta_2^2)$ was computed in Problem 2.44, and the first principal minor of the matrix $\mathbf{I}^{-1}(\theta)$ is θ_2^2. According to the asymptotic theory of m.l.e.'s, the maximal power for the given alternatives is $1 - F_1(\chi^2_{1-\alpha,1}; \lambda^2)$, where $\lambda^2 = \beta^2/\theta_2^2$.

3.74. In this case (see the solution to Problem 3.73)

$$\sup_{\Theta} L(\mathbf{x}; \theta) = L(\mathbf{x}; (\overline{x}, s)) = (2\pi e s^2)^{-n/2}, \; s^2 = S^2(\mathbf{x}),$$

$$\sup_{\Theta_0} L(\mathbf{x}; \theta) = L(\mathbf{x}; (\overline{x}, \theta_{20})) = (2\pi\theta_{20}^2)^{-n/2} e^{-ns^2/2\theta_{20}^2},$$

whence

$$\lambda_n(\mathbf{x}) = (t e^{-t+1})^{n/2}, \quad t = s^2/\theta_{20}^2.$$

Here the inequality $\lambda_n \leqslant c$, which defines the critical region of the likelihood ratio test, is written in the form $\{t \leqslant t_1\} \cup \{t \geqslant t_2\}$, $t_1 < t_2$, and the test is

$$\mathscr{X}_1^* = \{\mathbf{x}: S^2(\mathbf{x})/\theta_{20}^2 \leqslant t_1 \quad \text{or} \quad S^2(\mathbf{x})/\theta_{20}^2 \geqslant t_2\}.$$

By the hint, the power function of this test is

$$W(\theta) = F_{n-1}\left(\frac{n\theta_{20}^2}{\theta_2^2} t_1\right) + 1 - F_{n-1}\left(\frac{n\theta_{20}^2}{\theta_2^2} t_2\right).$$

Choosing $t_1 = \chi^2_{\alpha_1, n-1}/n$, $t_2 = \chi^2_{1-\alpha_2, n-1}/n$ for $\alpha_1 + \alpha_2 = \alpha$, we find that $W(\theta_0) = \alpha$, q.e.d. (compare to Problem 3.65 (6)).

For the test to be unbiased, the condition $W'(\theta_0) = 0$ should be met and we must solve the equation

$$\chi^2_{\alpha_1, n-1} k_{n-1}(\chi^2_{\alpha_1, n-1}) = \chi^2_{1-\alpha_2, n-1} k_{n-1}(\chi^2_{1-\alpha_2, n-1}).$$

Remark. The asymptotic (as $n \to \infty$) version of this test was investigated in [7, pp. 212-213].

3.75. Since the maximum likelihood estimate for the parameter θ in the model $Bi(1, \theta)$ constructed from the sample of size n is $\hat{\theta}_n = \overline{X}$ (see Problems 2.84 and 2.48), we have the statistic

$$\lambda_n = \frac{\theta_0^{n\overline{X}}(1-\theta_0)^{n(1-\overline{X})}}{\overline{X}^{n\overline{X}}(1-\overline{X})^{n(1-\overline{X})}}.$$

In our case the model is polynomial with $N = 2$ outcomes, and therefore (see [7, pp. 207-208]) as $n \to \infty$ the limiting distributions of the statistics $-2\ln\lambda_n$ and

$$X_n^2 = \frac{(T - n\theta_0)^2}{n\theta_0} + \frac{(n - T - n(1 - \theta_0))^2}{n(1 - \theta_0)} = \frac{(T - n\theta_0)^2}{n\theta_0(1 - \theta_0)}, \quad T = n\overline{X},$$

coincide under the hypothesis H_0 and are $\chi^2(1)$. This means that $\mathcal{L}_{\theta_0}(T) \sim \mathcal{N}(n\theta_0, n\theta_0(1 - \theta_0))$ as $n \to \infty$. Thus the $\{\lambda_n \leqslant c\}$-test is asymptotically equivalent to the $\{X_n^2 \geqslant t\}$-test, which coincides with the test constructed in Problem 3.63 for $k = 1$.

3.76. In our case the m.l.e. is $\hat{\theta}_n = \overline{X} = T/n$ (see Problem 2.109), and we have $\lambda_n = (n\theta_0/T)^T e^{T - n\theta_0}$. It follows from the solution to Problem 3.64 that the statistic $Q_n^{(2)} = (T - n\theta_0)^2/n\theta_0$, and the likelihood ratio test is asymptotically equivalent to the $\{|T - n\theta_0|/\sqrt{n\theta_0} \geqslant t\}$-test investigated in Problem 3.64.

3.77. We use $L_j(\theta_j) = \theta_j^{n_j\overline{X}_j}(1 - \theta_j)^{n_j(1 - \overline{X}_j)}$ to denote the likelihood function for the jth sample, $j = 1, \ldots, k$. Since the samples are independent, the likelihood function for all the data is $L(\theta_1, \ldots, \theta_k) = L_1(\theta_1)\ldots L_k(\theta_k)$. The m.l.e. of the Bernoulli model parameter coincides with the arithmetic mean of the sample, and we have

$$\max_{\theta_1, \ldots, \theta_k} L(\theta_1, \ldots, \theta_k) = \prod_{j=1}^{k} \max_{\theta_j} L_j(\theta_j) = \prod_{j=1}^{k} L_j(\overline{X}_j)$$
$$= \prod_{j=1}^{k} \overline{X}_j^{n_j\overline{X}_j}(1 - \overline{X}_j)^{n_j(1 - \overline{X}_j)},$$

$$\max_{\theta_1 = \ldots = \theta_k} L(\theta_1, \ldots, \theta_k) = \max_{\theta} L(\theta, \ldots, \theta) = \max_{\theta} \theta^{n\overline{X}}(1 - \theta)^{n(1 - \overline{X})}$$
$$= \overline{X}^{n\overline{X}}(1 - \overline{X})^{n(1 - \overline{X})},$$

where $\overline{X} = \frac{1}{n}(n_1\overline{X}_1 + \ldots + n_k\overline{X}_k)$, $n = n_1 + \ldots + n_k$.

Then the likelihood ratio statistic has the form

$$\lambda_{n_1\ldots n_k} = \overline{X}^{n\overline{X}}(1 - \overline{X})^{n(1 - \overline{X})} \Big/ \prod_{j=1}^{k} \overline{X}_j^{n_j\overline{X}_j}(1 - \overline{X}_j)^{n_j(1 - \overline{X}_j)}$$

$$= \prod_{j=1}^{k} \left(\frac{\overline{X}}{\overline{X}_j}\right)^{n_j\overline{X}_j} \left(\frac{1 - \overline{X}}{1 - \overline{X}_j}\right)^{n_j(1 - \overline{X}_j)}.$$

The dimensionality of the null hypothesis is dim $H_0 = 1$ (one degree of freedom), and the asymptotic likelihood ratio test has the form

$$\mathscr{X}_{1\alpha} = \left\{ 2 \sum_{j=1}^{k} n_j[\bar{x}_j(\ln \bar{x}_j - \ln \bar{x}) + (1 - \bar{x}_j) \right.$$
$$\left. \times (\ln (1 - \bar{x}_j) - \ln (1 - \bar{x}))] \geqslant \chi^2_{1-\alpha,k-1} \right\}.$$

The standard χ^2 uniformity test [7, p. 161] is

$$\left\{ \frac{1}{\bar{x}(1 - \bar{x})} \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2 \geqslant \chi^2_{1-\alpha,k-1} \right\}.$$

3.78. We solve this problem as the previous one. Using the same notations and taking into account that

$$L_j(\theta_j) = e^{-n_j\theta_j}\theta_j^{n_j\bar{X}_j} \Big/ \prod_{r=1}^{n_j} X_{jr}!, \quad j = 1, \ldots, k,$$

we find that

$$\lambda_{n_1 \ldots n_k} = \overline{X}^{n\overline{X}} \Big/ \prod_{j=1}^{k} \overline{X}_j^{n_j\overline{X}_j} = \prod_{j=1}^{k} \left(\frac{\overline{X}}{\overline{X}_j} \right)^{n_j\overline{X}_j}.$$

For $n_1, \ldots, n_k \to \infty$ the likelihood ratio test has the form

$$\mathscr{X}_{1\alpha} = \left\{ 2 \sum_{j=1}^{k} n_j\bar{x}_j(\ln \bar{x}_j - \ln \bar{x}) \geqslant \chi^2_{1-\alpha,k-1} \right\}.$$

3.79. We use $L_j(\theta_{1j}, \theta_2)$ to denote the likelihood function for the jth sample, $j = 1, \ldots, k$, and $L(\theta_{11}, \ldots, \theta_{1k}, \theta_2) = \prod_{j=1}^{k} L_j(\theta_{1j}, \theta_2)$ to denote the likelihood function for all the data. We stipulate that $n = n_1 + \ldots + n_k$, $S_0^2 = \frac{1}{n} \sum_{j=1}^{k} n_j S_j^2$, $\overline{X} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{X}_j$. Let S^2 be the sample variance for all the data. Then (see Problem 2.114) $(\overline{X}_1, \ldots, \overline{X}_k, S_0)$ and $L(\overline{X}_1, \ldots, \overline{X}_k, S_0) = (2\pi e S_0^2)^{-n/2}$ are m.l.e.'s for the parameters $(\theta_{11}, \ldots, \theta_{1k}, \theta_2)$. Under the hypothesis H_0 all the data may be considered to be a sample of size n from the population $\mathcal{N}(\theta_1, \theta_2^2)$. Therefore (see Problem 2.86),

$$\max_{\theta_1, \theta_2} L(\theta_1, \ldots, \theta_1, \theta_2) = L(\overline{X}, \ldots, \overline{X}, S) = (2\pi e S^2)^{-n/2}.$$

Thus, the likelihood ratio statistic is

$$\lambda_{n_1,\ldots n_k} = \left(\frac{S^2}{S_0^2}\right)^{-n/2} = \left(nS^2 \Big/ \sum_{j=1}^{k} n_j S_j^2\right)^{-n/2}.$$

The number of the model's parameters is $k + 1$, and the dimensionality of the null hypothesis is 2. Therefore, the asymptotic likelihood ratio test has the form

$$\mathscr{X}_{1\alpha} = \{n(\ln S^2 - \ln S_0^2) \geqslant \chi^2_{1-\alpha,k-1}\}.$$

If $k = 2$, then we can directly verify that

$$nS^2 = n_1 S_1^2 + n_2 S_2^2 + \frac{n_1 n_2}{n}(\overline{X}_1 - \overline{X}_2)^2,$$

whence

$$\lambda_{n_1 n_2} = (1 + T^2/(n-2))^{-n/2},$$

with T defined in the statement of the problem. Here the inequality $\lambda_{n_1 n_2} \leqslant c$ is equivalent in $|T| \geqslant t$, whence follows the required form of the test.

3.80. Let $L_j(\theta_{1j}, \theta_{2j})$ be the likelihood function for the jth sample, $j = 1, \ldots, k$, and $L = \prod_{j=1}^{k} L_j(\theta_{1j}, \theta_{2j})$ be the likelihood function for all the data. As in the previous problem, we have for the common model

$$\max L = \prod_{j=1}^{k} \max L_j(\theta_{1j}, \theta_{2j}) = \prod_{j=1}^{k} (2\pi e S_j^2)^{-n_j/2},$$

while under the hypothesis H_0

$$\max L = \max \prod_{j=1}^{k} L_j(\theta_{1j}, \theta_2) = (2\pi e S_0^2)^{-n/2}.$$

Whence

$$\lambda_{n_1,\ldots n_k} = \prod_{j=1}^{k} S_j^{n_j}/S_0^n = \prod_{j=1}^{k} (S_j/S_0)^{n_j}.$$

Here we have $2k$ parameters, and the dimensionality of the null hypothesis is $k + 1$. Therefore, the asymptotic likelihood ratio test for $n_1, \ldots, n_k \to \infty$ has the form

$$\mathscr{X}_{1\alpha} = \left\{ \sum_{j=1}^{k} n_j(\ln S_0^2 - \ln S_j^2) \geqslant \chi^2_{1-\alpha,k-1} \right\}.$$

For $k = 2$ we have

$$\lambda_{n_1 n_2} = \left(\frac{nS_1^2}{n_1 S_1^2 + n_2 S_2^2}\right)^{n_1/2} \left(\frac{nS_2^2}{n_1 S_1^2 + n_2 S_2^2}\right)^{n_2/2}$$

$$= \frac{n^{n/2}}{n_1^{n_1/2} n_2^{n_2/2}} \left(\frac{n_1 - 1}{n_2 - 1}\right)^{n_1/2} F^{n_1/2} \left(1 + \frac{n_1 - 1}{n_2 - 1} F\right)^{-n/2},$$

and the inequality $\lambda_{n_1 n_2} \leqslant c$ is equivalent to the condition $\{F \leqslant c_1\} \cup \{F \geqslant c_2\}$, $c_1 < c_2$. We use the hint and obtain the required test.

3.81. All X_i are distributed as $\mathcal{N}(\theta_1, \theta_2^2)$ under the hypothesis H_0. But in this case (see the solution to Problem 1.58) the distribution of the statistic $\eta = \dfrac{X_1 - \overline{X}}{\sqrt{n-1}\,S}$ is independent of the parameters (θ_1, θ_2) and is symmetric about zero with

$$\mathbf{P}(\eta > v) = \frac{1}{2} \mathrm{B}\left(1 - v^2; \frac{n-2}{2}, \frac{1}{2}\right), \quad 0 < v < 1.$$

Consequently,

$$\mathbf{P}(\mathscr{X}_{1\alpha} | H_0) = \mathrm{B}\left(1 - v_\alpha^2; \frac{n-2}{2}, \frac{1}{2}\right) = \alpha,$$

i.e., the probability to make an error when rejecting the hypothesis H_0 for $|\eta| > v_\alpha$ is α. Thus, $\mathscr{X}_{1\alpha}$ is a suitable critical region.

3.82. Since here the hypothesis H_0 is equivalent to $\operatorname{cov}(\xi_1, \xi_2) = 0$, by Problem 1.59 (c) we have

$$\mathscr{L}(\varrho_n \sqrt{(n-2)/(1-\varrho_n^2)} | H_0) = S(n-2).$$

Whence (due to the symmetry of Student's distribution)

$$\mathbf{P}(|\varrho_n| \sqrt{(n-2)/(1-\varrho_n^2)} > t_{1-\alpha/2, n-2} | H_0) = \alpha.$$

The latter inequality is equivalent to that defining the critical region $\mathscr{X}_{1\alpha}$, and hence $\mathbf{P}(\mathscr{X}_{1\alpha} | H_0) = \alpha$. This means that the probability to make an error when rejecting H_0 for $\varrho_n \in \mathscr{X}_{1\alpha}$ is α, i.e., $\mathscr{X}_{1\alpha}$ is the sought-for test.

3.83. By Problem 1.60, $\mathscr{L}(T_n | H_0) \to \mathcal{N}(0, 1)$ as $n \to \infty$ and, therefore, $\mathbf{P}(|T_n| \geqslant -u_{\alpha/2} | H_0) \to 2\Phi(u_{\alpha/2}) = \alpha$, q.e.d.

3.84. Suppose that the observable random variables $X_1, \ldots, X_n$ are independent and have the same and known mean μ and finite variances. If the hypothesis H_0 implies that all the variables X_i are similarly distributed, then for large n the statistic T_n is approximately normally distributed as $\mathcal{N}(0, 1)$ under H_0. Then we can use this statistic to construct the goodness of fit test for H_0 as in Problem 3.83.

3.85. (1) We have

$$W_n(\theta_0) = \mathbf{P}_{\theta_0}(T_n \geqslant \gamma_n) = \mathbf{P}_{\theta_0}(\sqrt{n}(T_n - \mu(\theta_0))/\sigma(\theta_0) \geqslant -u_\alpha),$$

and by property (a), $W_n(\theta_0) \to \Phi(u_\alpha) = \alpha$ as $n \to \infty$, i.e., the $\mathscr{S}_{1\alpha}$-test asymptotically has the significance level α.

(2) Reasoning as above, we have

$$W_n(\theta^{(n)}) = \mathbf{P}_{\theta^{(n)}}(T_n \geqslant \gamma_n) = \mathbf{P}_{\theta^{(n)}}(\sqrt{n}(T_n - \mu(\theta^{(n)}))/\sigma(\theta^{(n)}) \geqslant z_\alpha^{(n)}),$$

where

$$-z_\alpha(n) = \sqrt{n}(\mu(\theta^{(n)}) - \mu(\theta_0))/\sigma(\theta^{(n)}) + u_\alpha\sigma(\theta_0)/\sigma(\theta^{(n)}).$$

According to property (b), $-z_\alpha(n) \to \beta\mu'(\theta_0)/\sigma(\theta_0) + u_\alpha$ as $n \to \infty$. By property (a), these relations give the required limiting equality.

3.86. From the previous proof we have

$$\Phi(\beta\sqrt{e_1'} + u_\alpha) = \lim_{n\to\infty} W_n^{(1)}\left(\theta_0 + \frac{\beta}{\sqrt{n}}\right) = \lim_{n\to\infty} W_{N_n}^{(2)}\left(\theta_0 + \frac{\beta}{\sqrt{N_n}}\sqrt{\frac{N_n}{n}}\right)$$

$$= \lim_{n\to\infty} W_{N_n}^{(2)}\left(\theta_0 + \frac{\beta}{\sqrt{\lambda N_n}}\right) = \Phi(\beta\sqrt{e_2'/\lambda} + u_\alpha).$$

Whence $\sqrt{e_1'} = \sqrt{e_2'/\lambda}$, or $\lambda = e_2'/e_1'$.

3.87. In our case $\mathscr{L}_\theta(T_n^{(1)}) = \mathscr{N}(\theta, \sigma^2/n)$, and (see Problems 3.86 and 3.85) the test has the form

$$\mathscr{S}_{1\alpha}^{(1)} = \{\overline{X} \geqslant \theta_0 - u_\alpha\sigma/\sqrt{n}\}.$$

Its Pitman efficiency is

$$e_1(\beta, \alpha) = \Phi(\beta/\sigma + u_\alpha), \ e_1' = \sigma^{-2}.$$

We get $\mathscr{L}_\theta(T_n^{(2)}) \sim \mathscr{N}(\theta, \pi\sigma^2/(2n))$ in a similar way (see Problem 1.32), whence

$$\mathscr{S}_{1\alpha}^{(2)} = \left\{Z_{n,1/2} \geqslant \theta_0 - u_\alpha\sigma\sqrt{\frac{\pi}{2n}}\right\},$$

$$e_2(\beta, \alpha) = \Phi\left(\frac{\beta}{\sigma}\sqrt{\frac{2}{\pi}} + u_\alpha\right), \ e_2' = \frac{2}{\pi\sigma^2}.$$

Consequently, $\lambda = e_2'/e_1' = 2/\pi$.

3.88. We write the likelihood function $L(\mathbf{x}; \theta)$ as

$$L(\mathbf{x}; \theta) = ((2\pi)^n|\Sigma|)^{-1/2}\exp\left\{-\frac{1}{2}Q_\theta(\mathbf{x})\right\},$$

where the quadratic form $Q_\theta(\mathbf{x}) = (\mathbf{x} - \theta\mathbf{t})'\Sigma^{-1}(\mathbf{x} - \theta\mathbf{t})$ can be represented as a sum

$$Q_\theta(\mathbf{x}) = Q_0(\mathbf{x}) - 2\theta\mathbf{t}'\Sigma^{-1}\mathbf{x} + \theta^2 Q_0(\mathbf{t}).$$

We note that $\mathbf{1}$ is the last column (and the last row) of the matrix Σ, i.e., $(0\ldots01)\Sigma = \mathbf{1}'$, and find $\mathbf{t}'\Sigma^{-1} = (0\ldots01)$, whence $\mathbf{t}'\Sigma^{-1}\mathbf{x} = x_n$. Then

$$Q_\theta(\mathbf{x}) = Q_0(\mathbf{x}) + \theta^2 t_n - 2\theta x_n,$$

which means that $L(\mathbf{x};\ \theta) = g(T(\mathbf{x});\ \theta)h(\mathbf{x})$, where $T(\mathbf{x}) = x_n$ and $g(T;\ \theta) = \exp\left\{\theta T - \frac{1}{2}\theta^2 t_n\right\}$. According to the factorization test, X_n is a sufficient statistic, and all the statistical problems for this model are solved on the basis of this statistic. Since $\mathscr{L}_\theta(X_n) = \mathscr{N}(\theta t_n,\ t_n)$, we have a normal distribution with an unknown mean. Then our problem is equivalent to verifying a hypothesis on the unknown mean of a univariate normal distribution when one observation is given, which was made in Problems 3.47, 3.58, and 3.60.

To Chapter 4

4.1. We have

$$\hat{\beta} = \begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \begin{Vmatrix}\Sigma z_1^{(i)^2} & \Sigma z_1^{(i)}z_2^{(i)}\\ \Sigma z_1^{(i)}z_2^{(i)} & \Sigma z_2^{(i)^2}\end{Vmatrix}^{-1}\begin{pmatrix}\Sigma z_1^{(i)}X_i\\ \Sigma z_2^{(i)}X_i\end{pmatrix}$$

$$= \frac{1}{\Sigma z_1^{(i)^2}\Sigma z_2^{(i)^2} - (\Sigma z_1^{(i)}z_2^{(i)})^2}\begin{pmatrix}\Sigma z_2^{(i)^2}\Sigma z_1^{(i)}X_i - \Sigma z_1^{(i)}z_2^{(i)}\Sigma z_2^{(i)}X_i\\ \Sigma z_1^{(i)^2}\Sigma z_2^{(i)}X_i - \Sigma z_1^{(i)}z_2^{(i)}\Sigma z_1^{(i)}X_i\end{pmatrix}.$$

4.2. If not all the t_i are the same, then

$$\hat{\beta}_1 = \overline{X} - \bar{t}\hat{\beta}_2,\ \hat{\beta}_2 = \sum_{i=1}^{n}(t_i - \bar{t})(X_i - \overline{X})\Big/\sum_{i=1}^{n}(t_i - \bar{t})^2.$$

Here $\mathbf{E}\hat{\beta}_i = \beta_i$, $i = 1, 2$, and

$$\mathbf{D}\hat{\beta}_1 = \frac{\sigma^2}{n}\sum_{i=1}^{n}t_i^2\Big/\sum_{i=1}^{n}(t_i - \bar{t})^2,\ \mathbf{D}\hat{\beta}_2 = \sigma^2\Big/\sum_{i=1}^{n}(t_i - \bar{t})^2.$$

Therefore, if $\sum_{i=1}^{n}(t_i - \bar{t})^2 \to \infty$ as $n \to \infty$, then the estimates $\hat{\beta}_i$ are consistent.

4.3. We have $\tilde{\sigma}^2 = S(\hat{\beta})/(n-2)$, where $S(\hat{\beta}) = \sum_{i=1}^{n}(X_i - \overline{X})^2 - \hat{\beta}_2^2\sum_{i=1}^{n}(t_i - \bar{t})^2$, and $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ are as in Problem 4.2. If $\mathbf{D}S(\hat{\beta}) = o(n^2)$ as $n \to \infty$, then $\tilde{\sigma}^2$ is a consistent estimate for σ^2. Specifically, for the normal

model we have $\mathbf{E}(S(\hat{\beta})/\sigma^2) = \chi^2(n-2)$ [7, p. 235] and, therefore, $\mathbf{D}S(\hat{\beta}) = 2\sigma^4(n-2) = o(n^2)$.

4.4. We have $\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\sigma^2\bar{t}\Big/\sum_{i=1}^{n}(t_i - \bar{t})^2$.

4.6. Here $I = \sum_{j=1}^{3} \beta_j \frac{(b-a)^j}{j}$, $\hat{I} = \sum_{j=1}^{3} \hat{\beta}_j \frac{(b-a)^j}{j}$. Therefore, $\mathbf{E}\hat{I} = I$,

$$\mathbf{D}\hat{I} = \sum_{j,r=1}^{3} \frac{(b-a)^{j+r}}{jr} \operatorname{cov}(\hat{\beta}_j, \hat{\beta}_r).$$

4.9. The confidence intervals are found from

$$\left(\hat{\beta}_1 \pm t_{(1+\gamma)/2,n-2}\sqrt{S(\hat{\beta})\sum_{i=1}^{n} t_i^2\Big/\left[n(n-2)\sum_{i=1}^{n}(t_i-\bar{t})^2\right]}\right),$$

$$\left(\hat{\beta}_2 \pm t_{(1+\gamma)/2,n-2}\sqrt{S(\hat{\beta})\Big/\left[(n-2)\sum_{i=1}^{n}(t_i-\bar{t})^2\right]}\right),$$

$$S(\hat{\beta})/\chi^2_{(1+\gamma)/2,n-2} < \sigma^2 < S(\hat{\beta})/\chi^2_{(1-\gamma)/2,n-2}.$$

The γ-confidence ellipse has the form

$$\mathscr{G}_\gamma(\mathscr{A}) = \Big\{\beta: (\beta_1 - \hat{\beta}_1)^2 + 2\bar{t}(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + \frac{1}{n}\sum_{i=1}^{n} t_i^2(\beta_2 - \hat{\beta}_2)^2 < \frac{2}{n(n-2)} F_{\gamma,2,n-2}\Big\},$$

where $\hat{\beta}_1$, $\hat{\beta}_2$, and $S(\hat{\beta})$ are taken from Problems 4.2 and 4.3.

4.10. Since $\int_{-a}^{a} x(t)\,dt = 2a\beta_1$, we are dealing with the confidence interval for β_1 from the previous problem.

4.11. Here we have the theoretical dependence $a(t) = a(0) + vt$, and the confidence ellipse is constructed as in Problem 4.9.

4.14. We substitute in (4.11) $m = 3$, $\lambda_1 = (100)$, $\lambda_2 = (010)$, $\lambda_3 = (001)$, and find the sought-for system

$$\left\{\left(\hat{\beta}_j \pm \sqrt{\frac{3}{n-3} a^{jj} S(\hat{\beta}) F_{\gamma,3,n-3}}\right),\ \ j = 1, 2, 3\right\}.$$

4.15. We use $X_1, \ldots, X_8$ to denote the consecutive observations and obtain the following normal regression model:

$X_1 = \beta_1 + \varepsilon_1$, $X_2 = \beta_2 + \varepsilon_2$, $X_3 = \beta_1 + \beta_2 + \varepsilon_3$, $X_4 = 180 - \beta_2 - \beta_3 + \varepsilon_4$, $X_5 = \beta_3 + \varepsilon_5$, $X_6 = \beta_4 + \varepsilon_6$, $X_7 = \beta_3 + \beta_4 + \varepsilon_7$, $X_8 = 180 - \beta_1 - \beta_4 + \varepsilon_8$.

4.17. Taking into account the hint, we have

$$\mathbf{E}S(\hat{\beta}) = \mathbf{E}S(\beta) - \mathbf{E}(\hat{\beta} - \beta)'\mathbf{A}(\hat{\beta} - \beta),$$

where $\mathbf{E}S(\beta) = \mathbf{E}\sum_{i=1}^{n} \varepsilon_i^2 = n\sigma^2$ (see (4.4)) and

$$\mathbf{E}(\hat{\beta} - \beta)'\mathbf{A}(\hat{\beta} - \beta) = \sum_{i,j=1}^{k} a_{ij} \operatorname{cov}(\hat{\beta}_i, \hat{\beta}_j)$$
$$= \sigma^2 \sum_{i,j=1}^{k} a_{ij}a^{ij} = \sigma^2 \operatorname{tr}(\mathbf{A}\mathbf{A}^{-1}) = k\sigma^2.$$

We then obtain $\mathbf{E}S(\hat{\beta}) = (n - k)\sigma^2$. Since $\mathbf{X} - \mathbf{Z}'\hat{\beta} = \mathbf{B}\mathbf{X}$ (see (4.5)) and $\mathbf{B}^2 = \mathbf{B}$, we have $S(\hat{\beta}) = \mathbf{X}'\mathbf{B}^2\mathbf{X} = \mathbf{X}'\mathbf{B}\mathbf{X}$.

4.18. Here $\mathbf{A} = \mathbf{Z}\mathbf{Z}'$ is a diagonal matrix with the diagonal elements $\mathbf{z}_j'\mathbf{z}_j$, $j = 1, \ldots, k$, where $\mathbf{z}_j$ is the jth column of the matrix $\mathbf{Z}'$. Therefore, $\hat{\beta}_j = \mathbf{z}_j'\mathbf{X}/\mathbf{z}_j'\mathbf{z}_j$, $\mathbf{D}\hat{\beta}_j = \sigma^2/\mathbf{z}_j'\mathbf{z}_j$, $\operatorname{cov}(\hat{\beta}_j, \hat{\beta}_r) = 0$, $j \neq r$.

4.21. The likelihood function for the model (4.3) has the form (see (4.4))

$$L(\mathbf{x};\, \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} S(\mathbf{x};\, \beta)\right\},\ \theta = (\beta,\, \sigma^2),$$

and while maximizing it in β we minimize the quadratic form $S(\mathbf{x};\beta)$. It follows that the m.l.e. $\hat{\sigma}^2$ coincides with the l.s.e. $\hat{\beta}$ and the value $\hat{\sigma}^2$ of σ^2 minimizes $S(\hat{\beta})/\sigma^2 + n \ln \sigma^2$, whence $\hat{\sigma}^2 = S(\hat{\beta})/n$. Taking into account Problem 4.17, we get

$$\mathbf{E}\hat{\sigma}^2 - \sigma^2 = \left(\frac{n - k}{n} - 1\right)\sigma^2 = -\frac{k}{n}\sigma^2.$$

4.22. The solution follows from (4.10) for $m = 1$, $\mathbf{T} = \lambda'$. We should take into account that $F_{\gamma,1,n-k} = t^2_{(1+\gamma)/2,n-k}$ (see Problem 1.50).

4.23. Putting $\lambda = (1, t)$ in Problem 4.22 and using the results of Problems 4.2-4.4, we find the sought-for interval

$$\left(\overline{X} + (t - \overline{t})\hat{\beta}_2 \pm t_{(1+\gamma)/2,n-2}\sqrt{\frac{1}{n(n-2)} S(\hat{\beta})\left[1 + n(t - \overline{t})^2\Big/\sum_{i=1}^{n}(t_i - \overline{t})^2\right]}\right).$$

4.25. Substitute $\lambda_i = \mathbf{z}^{(i)}$, $i = 1, \ldots, n$, into (4.11).

4.27. We have $m = 1$ in (4.12) and

$$S_T = \min_{\beta_1} \sum_{i=1}^{n} (X_i - \beta_{20}t_i - \beta_1)^2 = S(\hat{\beta}) + (\hat{\beta}_2 - \beta_{20})^2 \sum_{i=1}^{n} (t_i - \bar{t})^2.$$

We also have $F_{1-\alpha,1,n-2} = t^2_{1-\alpha/2,n-2}$ (see the solution to Problem 4.22). We then find that the F-test (4.12) has the required form.

4.29. We have a normal regression model with $n = 8$, $\beta = (\mu_1, \mu_2, \mu_3, \mu_4)$, and the respective quadratic form is

$$S(\beta) = \sum_{i,j} (X_j^{(i)} - \mu_i)^2 = \sum_{i,j} (X_j^{(i)} - \overline{X^{(i)}})^2 + 2 \sum_i (\overline{X^{(i)}} - \mu_i)^2,$$

$$\overline{X^{(i)}} = \frac{1}{2} (X_1^{(i)} + X_2^{(i)}).$$

Hence the l.s.e. is $\hat{\mu}_i = \overline{X^{(i)}}$, $i = 1, \ldots, 4$, $\min S(\beta) = \sum_{i,j} (X_j^{(i)} - \overline{X^{(i)}})^2 \equiv S_1$, and $\hat{\sigma}^2 = S_1/4$.

Under the hypothesis H_0 we have

$$S_T = \min_{\mu} \sum_{i,j} (X_j^{(i)} - \mu)^2 = \min_{\mu} \left(\sum_{i,j} (X_j^{(i)} - \overline{X})^2 + n(\overline{X} - \mu)^2 \right)$$

$$= \sum_{i,j} (X_j^{(i)} - \overline{X})^2 = S_1 + 2 \sum_i (\overline{X^{(i)}} - \overline{X})^2,$$

and for $m = 3$ the F-test (4.12) has the form

$$\mathscr{I}_{1\alpha} = \left\{ \sum_i (\overline{X^{(i)}} - \overline{X})^2 / S_1 \geqslant \frac{3}{8} F_{1-\alpha,3,4} \right\}.$$

4.37. We use $\mathbf{Z}' = \|\mathbf{1}\alpha\|$ to denote an $n \times 2$ matrix of the column-vectors $\mathbf{1} = (1, \ldots, 1)$ and $\alpha = (\alpha_1, \ldots, \alpha_n)$ and find the solution in the form $\hat{\beta} = \mathbf{A}^{-1}\mathbf{Z}\mathbf{G}^{-1}\mathbf{Y}$, where $\mathbf{A} = \mathbf{Z}\mathbf{G}^{-1}\mathbf{Z}' = \left\| \begin{matrix} \mathbf{1}' \\ \hline \alpha' \end{matrix} \right\| \mathbf{G}^{-1}\|\mathbf{1}\alpha\|$. We then multiply the matrices ("row by column" with the block-matrices treated as the matrix elements) and reduce the result to the form $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2) = \dfrac{1}{\Delta} (\alpha'\mathbf{QY}, -\mathbf{1}'\mathbf{QY})$, where $\Delta = (\mathbf{1}'\mathbf{G}^{-1}\mathbf{1})(\alpha'\mathbf{G}^{-1}\alpha) - (\mathbf{1}'\mathbf{G}^{-1}\alpha)^2$ and $\mathbf{Q} = \mathbf{G}^{-1}(\alpha\mathbf{1}' - \mathbf{1}\alpha')\mathbf{G}^{-1}$. The matrix of the second moments will be

$$\mathbf{D}(\hat{\beta}) = \beta_2^2 \mathbf{A}^{-1} = \frac{\beta_2^2}{\Delta} \left\| \begin{matrix} \alpha'\mathbf{G}^{-1}\alpha & -\alpha'\mathbf{G}^{-1}\mathbf{1} \\ -\mathbf{1}'\mathbf{G}^{-1}\alpha & \mathbf{1}'\mathbf{G}^{-1}\mathbf{1} \end{matrix} \right\|.$$

To Chapter 5

5.1. (1) Here the sample space $\mathscr{X} = \{0, 1\}$ consists of two points, and we have two solutions for every $x \in \mathscr{X}$. We then have four decision functions $\delta_i = (\delta_i(0), \delta_i(1))$, $i = 1, 2, 3, 4$, $\delta_1 = (d_1, d_1)$, $\delta_2 = (d_1, d_2)$, $\delta_3 = (d_2, d_1)$, $\delta_4 = (d_2, d_2)$. Let $\mathbf{R}_i = (R(\theta_1, \delta_i), R(\theta_2, \delta_i))$ be the risk vector for the procedure δ_i, where

$$R(\theta, \delta_i) = L(\theta, \delta_i(0))(1 - \theta) + L(\theta, \delta_i(1))\theta.$$

Then

$$\mathbf{R}_1 = (0, 2),\ \mathbf{R}_2 = \left(\frac{1}{3}, \frac{2}{3}\right),\ \mathbf{R}_3 = \left(\frac{2}{3}, \frac{4}{3}\right),\ \mathbf{R}_4 = (1, 0).$$

The procedure δ_2 is preferable compared to δ_3, while $\delta_1, \delta_2, \delta_4$ are incomparable and hence form the set of admissible decision rules. Here $m(\delta_2) < m(\delta_4) < m(\delta_1)$, i.e., δ_2 is the minimax procedure.

(2) *The first solution.* For the Bayes risk $r(\delta_i) = \alpha R(\theta_1, \delta_i) + (1 - \alpha) \times R(\theta_2, \delta_i)$ we have $r(\delta_1) = 2(1 - \alpha)$, $r(\delta_2) = (2 - \alpha)/3$, $r(\delta_4) = \alpha$. If $\alpha \leqslant 1/2$, then $\min\limits_i r(\delta_i) = r(\delta_4)$, i.e., $\delta^* = \delta_4$. If $1/2 < \alpha \leqslant 4/5$, then $\min\limits_i r(\delta_i) = r(\delta_2)$, i.e., $\delta^* = \delta_2$. If $\alpha > 4/5$, then $\delta^* = \delta_1$. The Bayes risk is plotted in Fig. 8.

The second solution. We calculate the a posteriori distribution $\pi(\theta_i | x) = f(x; \theta_i)\pi(\theta_i)/f(x)$, $x = 0, 1$, $i = 1, 2$, where $f(x; \theta) = \theta^x(1 - \theta)^{1-x}$, $f(x) = f(x; \theta_1)\pi(\theta_1) + f(x; \theta_2)\pi(\theta_2) = \theta_1^x(1 - \theta_1)^{1-x}\alpha + \theta_2^x(1 - \theta_2)^{1-x} \times (1 - \alpha)$. We have $\pi(\theta_1|0) = \alpha(1 - \theta_1)/f(0)$, $\pi(\theta_2|0) = (1 - \alpha)(1 - \theta_2)/f(0)$, $\pi(\theta_1|1) = \alpha\theta_1/f(1)$, $\pi(\theta_2|1) = \dfrac{(1 - \alpha)\theta_2}{f(1)}$. In our case the average loss with respect to this a posteriori distribution is

$$\mathbf{E}(L(\theta, d_1)|0) = L(\theta_1, d_1)\pi(\theta_1|0) + L(\theta_2, d_1)\pi(\theta_2|0) = \frac{2(1 - \alpha)}{3f(0)}$$

for $x = 0$ and the solution $\delta(0) = d_1$, while the average loss for $\delta(0) = d_2$ is $2\alpha/3f(0)$. We compare the losses and see that for $\alpha \leqslant 1/2$ the loss for d_2 is

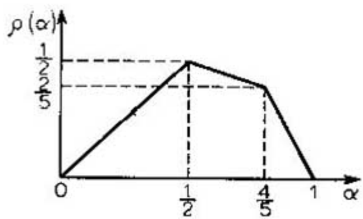


Fig. 8

smaller, i.e., $\delta^*(0) = d_2$, while for $\alpha > 1/2$ we have $\delta^*(0) = d_1$. We act in a similar way for the case of $x = 1$ to find that $\delta^*(1) = d_2$ for $\alpha \leqslant 4/5$, while $\delta^*(1) = d_1$ for $\alpha > 4/5$. We have thus found the Bayes solutions $\delta^*(x)$ at every point $x = 0, 1$ for all a priori distributions.

5.2. The average losses $\mathbf{E}(L(\theta, d)|x)$, calculated by the given formulas for the a priori distribution $\pi(\theta_1) = \alpha$, $\alpha \in [0, 1]$, are given in the following table:

	d_1	d_2	d_3
$x = 0$	$(1 - \alpha)/f(0)$	$3\alpha/4f(0)$	$(1 + 2\alpha)/8f(0)$
$x = 1$	$3(1 - \alpha)/f(1)$	$\alpha/4f(1)$	$(3 - 2\alpha)/8f(1)$

We compare these values to find the minimal one in every row (and thus find $\delta^*(x)$) and get $\delta^*(0) = d_2$ for $\alpha \leqslant 1/4$, $\delta^*(0) = d_3$ for $1/4 < \alpha \leqslant 7/10$, $\delta^*(0) = d_1$ for $\alpha > 7/10$, $\delta^*(1) = d_2$ for $\alpha \leqslant 3/4$, $\delta^*(1) = d_3$ for $3/4 < \alpha \leqslant 21/22$, $\delta^*(1) = d_1$ for $\alpha > 21/22$. Thus, the sought-for Bayes solutions $\delta^* = (\delta^*(0), \delta^*(1))$ are $\delta^* = (d_2, d_2)$ for $\alpha \leqslant 1/4$, $\delta^* = (d_3, d_2)$ for $1/4 < \alpha \leqslant 7/10$, $\delta^* = (d_1, d_2)$ for $7/10 < \alpha \leqslant 3/4$, $\delta^* = (d_1, d_3)$ for $3/4 < \alpha \leqslant 21/22$, $\delta^* = (d_1, d_1)$ for $\alpha > 21/22$. Since

$$\varrho(\alpha) = r(\delta^*) = \sum_{x=0}^{1} f(x)\mathbf{E}(L(\theta, \delta^*(x))|x),$$

we take the respective values of $\mathbf{E}(L(\theta, \delta^*(x))|x)$ from the table and find

$$\varrho(\alpha) = \begin{cases} \alpha & \text{for } 0 \leqslant \alpha \leqslant 1/4, \\ (1 + 4\alpha)/8 & \text{for } 1/4 < \alpha \leqslant 7/10, \\ (4 - 3\alpha)/4 & \text{for } 7/10 < \alpha \leqslant 3/4, \\ (11 - 10\alpha)/8 & \text{for } 3/4 < \alpha \leqslant 21/22, \\ 4(1 - \alpha) & \text{for } 21/22 < \alpha \leqslant 1. \end{cases}$$

The function $\varrho(\alpha)$ is plotted in Fig. 9.

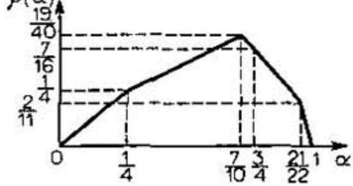


Fig. 9

5.3. For the decision functions of Problem 5.1 the risk vectors are

$$\mathbf{R}_1^{(1)} = (0,\ b),\ \mathbf{R}_2^{(1)} = \left(\frac{2a}{3}, \frac{b}{2}\right),\ \mathbf{R}_3^{(1)} = \left(\frac{a}{3}, \frac{b}{2}\right),\ \mathbf{R}_4^{(1)} = (a,\ 0)$$

in the first case and

$$\mathbf{R}_1^{(2)} = (0,\ b),\ \mathbf{R}_2^{(2)} = \left(\frac{3a}{4}, \frac{b}{2}\right),\ \mathbf{R}_3^{(2)} = \left(\frac{a}{4}, \frac{b}{2}\right),\ \mathbf{R}_4^{(2)} = (a,\ 0)$$

in the second case. We thus conclude that the procedure δ_3 is preferable compared to δ_2, and then δ_1, δ_3, and δ_4 are the admissible procedures. For the values $\alpha = \pi(\theta_1)$ with the Bayes solution $\delta^* = \delta_3$ the risks satisfy the inequality $r^{(2)}(\delta^*) < r^{(1)}(\delta^*)$.

5.4. Here the risk function is

$$R(\theta,\ \delta) = \sum_{k=0}^{3} L(\theta,\ \delta(k))C_3^k\theta^k(1 - \theta)^{3-k},$$

and the risk vectors $\mathbf{R}_i = (R(\theta_1,\ \delta_i),\ R(\theta_2,\ \delta_i))$ for the given procedures are $\mathbf{R}_1 = (5.94 \times 10^{-2},\ 0.792)$, $\mathbf{R}_2 = (5.96 \times 10^{-4},\ 0.972)$, $\mathbf{R}_3 = (2 \times 10^{-6},\ 0.999)$, respectively. The minimax solution is $\tilde{\delta} = \delta_1$ and $m(\tilde{\delta}) = 0.792$.

5.5. Here $f(x;\ \theta) = (1 - \theta)\theta^x$, $x = 0, 1, 2, \ldots$, and the risk function is

$$\begin{aligned} R(\theta,\ \delta_i) &= \sum_{x=0}^{\infty} L(\theta,\ \delta_i(x))f(x;\ \theta) \\ &= L(\theta,\ d_1)\sum_{x=0}^{i-1} f(x;\ \theta) + L(\theta,\ d_2)\sum_{x=i}^{\infty} f(x;\ \theta) \\ &= L(\theta,\ d_1)(1 - \theta^i) + L(\theta,\ d_2)\theta^i. \end{aligned}$$

The risk vector for the ith procedure is

$$\mathbf{R}_i = (R(\theta_1,\ \delta_i),\ R(\theta_2,\ \delta_i)) = (a\theta_1^i,\ b(1 - \theta_2^i)).$$

As i increases, the first coordinate decreases and the second increases. Therefore, for $a\theta_1 \leqslant b(1 - \theta_2)$ the maximal risk is $m(\delta_i) = b(1 - \theta_2^i)$, and hence the minimax procedure is $\tilde{\delta} = \delta_1$.

If $a\theta_1 > b(1 - \theta_2)$, then we find an integer i_0 from the conditions $a\theta_1^{i_0} \geqslant b(1 - \theta_2^{i_0})$, $a\theta_1^{i_0+1} < b(1 - \theta_2^{i_0+1})$ and obtain

$$m(\delta_i) = \begin{cases} a\theta_1^i, & i \leqslant i_0, \\ b(1 - \theta_2^i), & i > i_0. \end{cases}$$

In our case $\tilde{\delta} = \delta_{i_0}$ if $a\theta_1^{i_0} \leqslant b(1 - \theta_2^{i_0+1})$, otherwise $\tilde{\delta} = \delta_{i_0+1}$.

5.7. By the general theory (see Sec. 5.4) we have $h_1(x) = b\pi_2 f_2(x)$, $h_2(x) = a\pi_1 f_1(x)$, and therefore the Bayes solution $\delta^*(x)$ has the form

$$\delta^*(x) = \begin{cases} d_1 & \text{for } x \in W_1^*, \quad \text{where } W_1^* = \{x: h_1(x) \leqslant h_2(x)\} \\ d_2 & \text{for } x \in \overline{W}_1^*, \qquad\qquad\quad = \left\{x: \dfrac{f_2(x)}{f_1(x)} \leqslant \dfrac{a\pi_1}{b\pi_2}\right\}. \end{cases}$$

The respective risk vector is $(a\mathbf{P}_{\theta_1}(X \in \overline{W}_1^*), b\mathbf{P}_{\theta_2}(X \in W_1^*))$, and the Bayes risk is

$$r(\delta^*) = a\pi_1 \mathbf{P}_{\theta_1}(X \in \overline{W}_1^*) + b\pi_2 \mathbf{P}_{\theta_2}(X \in W_1^*).$$

For $\theta_1 < \theta_2$ in the indicated normal distributions we have $W_1^* = \left\{x: x \leqslant \dfrac{\theta_1 + \theta_2}{2} - \dfrac{\sigma^2 c}{\theta_2 - \theta_1}\right\}$, $c = \ln \dfrac{b\pi_2}{a\pi_1}$, and $\mathbf{P}_{\theta_2}(X \in W_1^*) = \Phi\left(-\dfrac{c + \varrho/2}{\sqrt{\varrho}}\right)$, $\mathbf{P}_{\theta_1}(X \in \overline{W}_1^*) = \Phi\left(\dfrac{c - \varrho/2}{\sqrt{\varrho}}\right)$, where $\varrho = (\theta_1 - \theta_2)^2/\sigma^2$. If $\theta_1 > \theta_2$, then the region W_1^* is defined by the inverted inequality and all other relations remain as before.

5.9. For the decision function $\delta(x) \equiv d$ the risk function is $R(\theta, \delta) = L(\theta, d)$, and the Bayes risk is

$$r(d) = \alpha L(0, d) + (1 - \alpha)L(1, d) = \alpha d^a + (1 - \alpha)(1 - d)^a.$$

Minimizing this with respect to d, we find the sought-for decision d^*, i.e., if $a = 1$, then $d^* = 0$ for $\alpha > 1/2$, $d^* = 1$ for $\alpha < 1/2$, while we may take any $d \in [0, 1]$ as d^* for $\alpha = 1/2$. If $a > 1$, we have

$$d^* = \left(1 + \left(\frac{\alpha}{1 - \alpha}\right)^{\frac{1}{a-1}}\right)^{-1}.$$

5.10. (1) We write $\mathbf{P}_{\theta_i}(X \in W_j^*) = \int\limits_{W_j^*} f_i(x)\, dx$ (for the absolutely continuous case) and use the definition of $h_j(x)$ to find

$$r(\delta^*) = \sum_{i=1}^{k} \pi_i R_i(\delta^*) = \sum_{j=1}^{k} \int\limits_{W_j^*} h_j(x)\, dx.$$

By the definition of the regions W_j^* the latter expression can be written as required.

(2) It is obvious that

$$\underline{l}\left(\sum_{i=1}^{k} \pi_i f_i(x) - \pi_j f_j(x)\right) \leqslant h_j(x) \leqslant \overline{l}\left(\sum_{i=1}^{k} \pi_i f_i(x) - \pi_j f_j(x)\right)$$

(we have taken into account that $l(j|j) = 0$, $j = 1, \ldots, k$). Then

$$\underline{l}\left(\sum_{i=1}^{k} p_i f_i(x) - \min_j \pi_j f_j(x)\right) \leqslant \min_j h_j(x) \leqslant \bar{l}\left(\sum_{i=1}^{k} \pi_i f_i(x) - \max_j \pi_j f_j(x)\right)$$

or (see the hint)

$$\underline{l} \sum_{i=2}^{k} \min\left(\pi_i f_i(x), \max_{j<i} \pi_j f_j(x)\right) \leqslant \min_j h_j(x)$$
$$\leqslant \bar{l} \sum_{i=2}^{k} \min\left(\pi_i f_i(x), \max_{j<i} \pi_j f_j(x)\right).$$

But

$$\min\left(\pi_i f_i(x), \max_{j<i} \pi_j f_j(x)\right) \leqslant \sum_{j<i} \min\left(\pi_i f_i(x), \pi_j f_j(x)\right),$$

which gives the upper estimate for $r(\delta^*)$ after the integration. We then have

$$\min\left(\pi_i f_i(x), \max_{j<i} \pi_j f_j(x)\right) \geqslant \min\left(\pi_i f_i(x), \pi_j f_j(x)\right),\ j = 1, \ldots, i-1,$$

and

$$\int \min\left(\pi_i f_i(x), \max_{j<i} \pi_j f_j(x)\right) dx \geqslant \max_{j<i} I_{ij},$$

which gives the lower estimate for $r(\delta^*)$.

The estimates become strict equalities for $k = 2$ and $l(2|1) = l(1|2) = 1$.

5.11. We have

$$I_{12} = \pi_1 \int_{\mathscr{X}_1} f_1(\mathbf{x})\, d\mathbf{x} + \pi_2 \int_{\mathscr{X}_1} f_2(\mathbf{x})\, d\mathbf{x},\ \mathbf{x} = (x_1, \ldots, x_r),$$

where $\mathscr{X}_1 = \{\mathbf{x}: \pi_1 f_1(\mathbf{x}) \leqslant \pi_2 f_2(\mathbf{x})\}$ and

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{r/2}\sqrt{|\mathbf{A}|}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu^{(i)})'\mathbf{A}^{-1}(\mathbf{x} - \mu^{(i)})\right\},\ i = 1, 2.$$

Simple transformations allow us to write $\mathscr{X}_1$ as

$$\mathscr{X}_1 = \left\{\mathbf{x}: \mathbf{a}'\mathbf{x} - \frac{1}{2}\mathbf{a}'(\mu^{(1)} + \mu^{(2)}) \leqslant \ln\frac{\pi_2}{\pi_1}\right\},\ \mathbf{a} = \mathbf{A}^{-1}(\mu^{(1)} - \mu^{(2)}).$$

Consider the random variable $Y = \mathbf{a}'\mathbf{X} - \dfrac{1}{2}\mathbf{a}'(\mu^{(1)} + \mu^{(2)})$. If $\mathscr{L}(\mathbf{X}) = \mathscr{N}(\mu^{(1)}, \mathbf{A})$, then $\mathscr{L}(Y) = \mathscr{N}(\varrho/2, \varrho)$ and

$$\int_{\mathscr{X}_1} f_1(\mathbf{x})\, d\mathbf{x} = \mathbf{P}\left(Y \leqslant \ln\frac{\pi_2}{\pi_1}\right) = \Phi\left(\left(\ln\frac{\pi_2}{\pi_1} - \frac{\varrho}{2}\right)\Big/\sqrt{\varrho}\right).$$

We find

$$\int_{\overline{\mathscr{W}_1}} f_2(\mathbf{x})\,d\mathbf{x} = 1 - \Phi\left(\left(\ln\frac{\pi_2}{\pi_1} + \frac{\varrho}{2}\right)\Big/\sqrt{\varrho}\right) = \Phi\left(\left(\ln\frac{\pi_1}{\pi_2} - \frac{\varrho}{2}\right)\Big/\sqrt{\varrho}\right)$$

in a similar way, q.e.d. Specifically, for $\pi_1 = \pi_2 = 1/2$ we have $I_{12} = \Phi\left(-\frac{\sqrt{\varrho}}{2}\right)$.

For $\lambda_1 > \lambda_2$ in Poisson's distributions $\Pi(\lambda_1)$ and $\Pi(\lambda_2)$ we have

$$I_{12} = \pi_1 e^{-\lambda_1}\sum_{r\in\mathscr{W}_1}\frac{\lambda_1^r}{r!} + \pi_2 e^{-\lambda_2}\sum_{r\in\overline{\mathscr{W}_1}}\frac{\lambda_2^r}{r!},$$

where

$$\mathscr{W}_1 = \left\{r:\ \pi_1 e^{-\lambda_1}\frac{\lambda_1^r}{r!} \leqslant \pi_2 e^{-\lambda_2}\frac{\lambda_2^r}{r!}\right\} = \left\{r:\ r \leqslant r_0 = \left[\frac{\lambda_1 - \lambda_2 + \ln\dfrac{\pi_2}{\pi_1}}{\ln\dfrac{\lambda_1}{\lambda_2}}\right]\right\},$$

$[\cdot]$ being the integer part. Thus,

$$I_{12} = \pi_1\Pi(r_0;\ \lambda_1) + \pi_2(1 - \Pi(r_0;\ \lambda_2)),$$

where $\Pi(r;\ \lambda) = \sum_{i=0}^{r} e^{-\lambda}\frac{\lambda^i}{i!}$.

5.12. We use the results and notations of the previous problem and find that the best classification region has the form

$$W_1^* = \{\mathbf{x}:\ h_1(\mathbf{x}) = l(1|2)\pi_2 f_2(\mathbf{x}) \leqslant h_2(\mathbf{x}) = l(2|1)\pi_1 f_1(\mathbf{x})\}$$

$$= \left\{\mathbf{x}:\ \mathbf{a}'\mathbf{x} - \frac{1}{2}\mathbf{a}'(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)}) \geqslant c = \ln\frac{\pi_2 l(1|2)}{\pi_1 l(2|1)}\right\}.$$

Thus, in the Bayes solution δ^* we assume that the distribution $\mathscr{N}(\boldsymbol{\mu}^{(1)}, \mathbf{A})$ is true for the observation $\mathbf{x} \in W_1^*$, otherwise (for $\mathbf{x} \in W_2^* = \overline{W_1^*}$) we assume that $\mathscr{N}(\boldsymbol{\mu}^{(2)}, \mathbf{A})$ is the true distribution. The risk vector (see the above solution) is

$$\mathbf{R}(\delta^*) = \left(l(2|1)\int_{W_2^*} f_1(\mathbf{x})\,d\mathbf{x},\ l(1|2)\int_{W_1^*} f_2(\mathbf{x})\,d\mathbf{x}\right)$$

$$= \left(l(2|1)\Phi\left(\frac{c - \varrho/2}{\sqrt{\varrho}}\right),\ l(1|2)\Phi\left(-\frac{c + \varrho/2}{\sqrt{\varrho}}\right)\right).$$

Consequently, the Bayes risk is

$$r(\delta^*) = \pi_1 l(2|1)\Phi\left(\frac{c - \varrho/2}{\sqrt{\varrho}}\right) + \pi_2 l(1|2)\Phi\left(-\frac{c + \varrho/2}{\sqrt{\varrho}}\right).$$

For $l(2|1) = l(1|2) = 1$, $\pi_1 = \pi_2 = 1/2$ we have $c = 0$ and $r(\delta^*) = \Phi\left(-\frac{\sqrt{\varrho}}{2}\right)$. In order to obtain the minimax solution $\bar{\delta}$, we find c (and the least favourable a priori distribution (π_1, π_2)) from the equation

$$l(2|1)\Phi\left(\frac{c - \varrho/2}{\sqrt{\varrho}}\right) = l(1|2)\Phi\left(-\frac{c + \varrho/2}{\sqrt{\varrho}}\right).$$

For $l(2|1) = l(1|2) = 1$ the solution is $c = 0$ (i.e., $\pi_1 = \pi_2 = 1/2$), and then the maximal risk is $m(\bar{\delta}) = \Phi\left(-\frac{\sqrt{\varrho}}{2}\right)$. The minimax rule is defined by the region W_1^* for $c = 0$.

5.13. (1) If $\mathscr{L}(\xi) = Bi(m, \theta)$, then $f(\mathbf{x}; \theta) = \theta^{\Sigma x_i}(1 - \theta)^{\Sigma(m - x_i)} = \theta^x(1 - \theta)^{nm - x}$, and the a priori distribution density of the parameter is $\pi(\theta) \simeq \theta^{a-1}(1 - \theta)^{b-1}$. For the a posteriori density we have

$$\pi(\theta|\mathbf{x}) \simeq \pi(\theta)f(\mathbf{x}; \theta) \simeq \theta^{a+x-1}(1 - \theta)^{b+nm-x-1},$$

i.e., $\pi(\theta|\mathbf{x})$ is the density of the beta distribution $B(a + x, b + nm - x)$.

(2) Here $f(\mathbf{x}; \theta) \simeq \theta^{\Sigma x_i}(1 - \theta)^{nr}$, $\pi(\theta)$ is given above, consequently,

$$\pi(\theta|\mathbf{x}) \simeq \theta^{a+x-1}(1 - \theta)^{b+nr-1}.$$

(3) Here $f(\mathbf{x}; \theta) = e^{-n\theta}\theta^{\Sigma x_i}$, $\pi(\theta) = \theta^{\lambda-1}e^{-\theta/a}$, and hence

$$\pi(\theta|\mathbf{x}) \simeq \theta^{\lambda+x-1}e^{-\theta(na+1)/a}$$

is the density of the gamma distribution $\Gamma\left(\frac{a}{na + 1}, \lambda + x\right)$.

(4) Here $f(\mathbf{x}; \theta) = \theta^n e^{-\theta\Sigma x_i}$, $\pi(\theta)$ is given above, consequently,

$$\pi(\theta|\mathbf{x}) \simeq \theta^{\lambda+n-1}e^{-\theta(ax+1)/a}.$$

(5) Using the indicator, we write the sample density in the form $f(\mathbf{x}; \theta) = \theta^{-n}I(\theta \geqslant x_{(n)})$, $x_{(n)} = \max(x_1, \ldots, x_n)$. Similarly, $\pi(\theta) \simeq \theta^{-\alpha-1}I(\theta \geqslant a)$, whence

$$\pi(\theta|\mathbf{x}) \simeq \theta^{-\alpha-n-1}I(\theta \geqslant \max(a, x_{(n)})).$$

(6) If $h_1, \ldots, h_N$ are non-negative integers which meet the condition $h_1 + \ldots + h_N = n$, then $f(\mathbf{h}; \theta) = \theta_1^{h_1}\ldots\theta_N^{h_N}$, and hence

$$\pi(\theta|\mathbf{h}) \simeq \theta_1^{\alpha_1+h_1-1}\ldots\theta_N^{\alpha_N+h_N-1},$$

i.e., we obtain the Dirichlet $D(\alpha + \mathbf{h})$ distribution density.

(7) The distribution density is

$$f(\mathbf{x};\,\theta) \asymp \exp\left\{-\frac{1}{2b^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}$$

$$= \exp\left\{-\frac{n}{2b^2}(\theta-\overline{x})^2 - \frac{1}{2b^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\},$$

and the a priori density is $\pi(\theta) \asymp \exp\left\{-\frac{1}{2\sigma^2}(\theta-\mu)^2\right\}$. Then

$$\pi(\theta|\mathbf{x}) \asymp \pi(\theta)f(\mathbf{x};\,\theta) \asymp \exp\left\{-\frac{1}{2\sigma^2}(\theta-\mu)^2 - \frac{n}{2b^2}(\theta-\overline{x})^2\right\}$$

$$\asymp \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{n}{b^2}\right)\theta^2 + \theta\left(\frac{\mu}{\sigma^2}+\frac{x}{b^2}\right)\right\},$$

where we have omitted the powers independent of θ. The latter expression is proportional to $\exp\left\{-\frac{1}{2\sigma_1^2}(\theta-\mu_1)^2\right\}$, q.e.d.

5.14. Let $X = x \in \{1, \ldots, n-1\}$. The a posteriori distribution is $\pi(\theta|x) \asymp \theta^x(1-\theta)^{n-x}$, and the average loss for the decision $\delta(x) = d$ with respect to this distribution is proportional to

$$\int_0^1 (d-\theta)^2\theta^{x-1}(1-\theta)^{n-x-1}\,d\theta = d^2\mathrm{B}(x,\,n-x) - 2d\mathrm{B}(x+1,\,n-x)$$

$$+\,\mathrm{B}(x+2,\,n-x) = c_1\left(d-\frac{x}{n}\right)^2 + c_2.$$

This expression attains minimum for $d = x/n$. If $x = 0$ or n, then the integral is finite only for $d = 0$ (or $d = 1$). Thus, $\delta^*(x) = x/n$ for any x. We also have

$$R(\theta,\,\delta^*) = \mathbf{E}_\theta\left(\frac{X}{n}-\theta\right)^2\Big/\theta(1-\theta) = \mathbf{D}_\theta\left(\frac{X}{n}\right)\Big/\theta(1-\theta) = \frac{1}{n} \equiv \text{const.}$$

Consequently, δ^* is the minimax solution with the risk $r(\delta^*) = 1/n$.

5.15. By virtue of Problem 5.13 (1), the a posteriori density is $\pi(\theta|x) \asymp \theta^{a+x-1}(1-\theta)^{b+n-x-1}$, and the average loss with respect to this distribution is

$$\int_0^1 (d-\theta)^2\theta^{a+x-1}(1-\theta)^{b+n-x-1}\,d\theta = c_1\left(d-\frac{x+a}{n+a+b}\right)^2 + c_2.$$

It follows that $\delta^*(x) = \dfrac{x + a}{n + a + b}$. We calculate the risk function

$$R(\theta, \delta^*) = \mathbf{E}_\theta \left(\frac{X + a}{n + a + b} - \theta\right)^2 = \mathbf{D}_\theta \left(\frac{X + a}{n + a + b}\right) + \left(\frac{n\theta + a}{n + a + b} - \theta\right)^2$$
$$= \frac{(a - \theta(a + b))^2 + n\theta(1 - \theta)}{(n + a + b)^2}.$$

The condition $R(\theta, \delta^*) \equiv$ const is met for $a = b = \sqrt{n}/2$. Consequently, the minimax solution is

$$\bar{\delta}(x) = \frac{x + \sqrt{n}/2}{n + \sqrt{n}},$$

and its risk is

$$m(\bar{\delta}) = R(\theta, \bar{\delta}) = \frac{1}{4(1 + \sqrt{n})^2}.$$

5.16. We write the average loss for the decision $\delta(x) = d$ with respect to the a posteriori distribution of $\pi(\theta|x)$ in the form

$$\mathbf{E}((\theta - d)^2|x) = \mathbf{E}((\theta - \delta^*(x) + \delta^*(x) - d)^2|x)$$
$$= \mathbf{D}(\theta|x) + (\delta^*(x) - d)^2 \geqslant \mathbf{D}(\theta|x).$$

The equality only holds for $d = \delta^*(x)$, and therefore $\delta^*(x)$ is the sought-for decision with the conditional (for $X = x$) risk $\mathbf{D}(\theta|x)$. Then the Bayes risk has the indicated form.

In Problem 5.15 the a posteriori mean of the parameter is equal to the first moment of the distribution $B(a + x, b + n - x)$, i.e.,

$$\delta^*(x) = \mathbf{E}(\theta|x) = \frac{a + x}{n + a + b}.$$

5.17. Here $\mathscr{L}_\theta(X) = \overline{Bi}(r, \theta)$. By Problem 5.13 (2) the a posteriori distribution is $\mathscr{L}(\theta|x) = B(a + x, b + r)$. Using the formulas for the moments of the beta distribution (see the introduction to Chap. 1), we find from Problem 5.16 that the sought-for Bayes estimate is

$$\delta^*(x) = \mathbf{E}(\theta|x) = \frac{a + x}{a + b + r + x}.$$

5.18. By Problem 5.13 (3) the a posteriori distribution is

$$\mathscr{L}(\theta|\mathbf{x}) = \Gamma\left(\frac{a}{na + 1}, \lambda + x\right), \quad \mathbf{x} = (x_1, \ldots, x_n),\ x = \sum_{i=1}^{n} x_i.$$

Using the formulas for the moments of the gamma distribution (see the introduction to Chap. 1), we find from Problem 5.16 that

$$\delta^*(\mathbf{x}) = \mathbf{E}(\theta|\mathbf{x}) = \frac{a(\lambda + x)}{na + 1}, \ \mathbf{D}(\theta|\mathbf{x}) = \frac{a^2(\lambda + x)}{(na + 1)^2}.$$

Since $\mathscr{L}_\theta(X) = \Pi(n\theta)$ (see Problem 1.39 (4)), the formula for the total expectation gives $\mathbf{E}X = \mathbf{E}(\mathbf{E}_\theta(X)) = \mathbf{E}(n\theta) = na\lambda$ and, therefore,

$$r(\delta^*) = \mathbf{E}\mathbf{D}(\theta|X) = \frac{a^2}{(na + 1)^2}(\lambda + \mathbf{E}X) = \frac{\lambda a^2}{na + 1}.$$

We minimize the quantity $\dfrac{\lambda a^2}{na + 1} + cn$ in n and obtain the optimal number of observations

$$n^* = \left(\frac{\lambda a}{c}\right)^{1/2} - \frac{1}{a}.$$

Remark. The number n^* must be a non-negative integer. If we obtain a negative number, we set $n^* = 0$ (i.e., the observations are not needed). In other cases we take the smallest integer which is greater than or equal to the value obtained for the required observation number. This is true for all similar problems.

5.20. By Problem 5.13 (4) the value d^* of the Bayes estimate for $\mathbf{X} = \mathbf{x}$ is found by minimizing

$$\mathbf{E}(L(\theta, d)|\mathbf{x}) = \frac{1}{\Gamma(\lambda + n)} \int_0^\infty \left(d - \frac{ax + 1}{a\theta}\right)^2 \theta^{\lambda + n - 1} e^{-\theta}\, d\theta$$

in d. We solve $\dfrac{\partial}{\partial d}\mathbf{E}(L(\theta, d)|\mathbf{x}) = 0$ with respect to d and find the required expression for δ^*.

Since $\mathscr{L}_\theta\left(\sum_{i=1}^{n} X_i\right) = \Gamma(1/\theta, n)$ (see Problem 1.39 (2)), we have

$$\mathbf{E}_\theta\delta^* = \frac{\theta + an}{\theta a(\lambda + n - 1)}, \quad \mathbf{D}_\theta\delta^* = \frac{n}{\theta^2(\lambda + n - 1)^2}.$$

The risk function is

$$R(\theta, \delta^*) = \mathbf{E}_\theta\left(\delta^* - \frac{1}{\theta}\right)^2 = \mathbf{D}_\theta\delta^* + \left(\mathbf{E}_\theta\delta^* - \frac{1}{\theta}\right)^2$$

$$= \frac{a^2 n + (\theta - a(\lambda - 1))^2}{\theta^2 a^2(\lambda + n - 1)^2}.$$

Finally, $r(\delta^*) = \mathbf{E}R(\theta, \delta^*)$, and we use the formulas for the moments of the gamma distribution to find the required result. The number n^* is found by minimizing $r(\delta^*) + cn$ in n.

5.21. The mean and variance of Pareto's distribution with the parameters a and $\alpha > 2$ are $\dfrac{\alpha a}{\alpha - 1}$ and $\dfrac{\alpha a^2}{(\alpha - 1)^2(\alpha - 2)}$, respectively. Whence (see Problem 5.13 (5)) we find the required δ^* using Problem 5.16. Besides,

$$\mathbf{D}(\theta | \mathbf{x}) = \frac{n + \alpha}{(n + \alpha - 1)^2(n + \alpha - 2)} (\max (a, x_{(n)}))^2.$$

In order to find the risk $r(\delta^*) = \mathbf{E}\mathbf{D}(\theta | \mathbf{X})$ it is sufficient to calculate

$$\mathbf{E}(\max (a, X_{(n)}))^2 = \mathbf{E}(\mathbf{E}_\theta(\max (a, X_{(n)}))^2).$$

We write

$$(\max (a, X_{(n)}))^2 = a^2 I(X_{(n)} \leqslant a) + X_{(n)}^2 I(X_{(n)} > a),$$

where $I(A)$ is the indicator of the event A. Since the distribution density $X_{(n)}$ for the given θ is nt^{n-1}/θ^n, $0 \leqslant t \leqslant \theta$ (see Problem 1.35), we have

$$\mathbf{E}_\theta(\max (a, X_{(n)}))^2 = \frac{na^2}{\theta^n} \int_0^a t^{n-1}\, dt + \frac{n}{\theta^n} \int_a^\theta t^{n+1}\, dt$$

$$= \frac{n}{n + 2} \theta^2 + \frac{2a^{n+2}}{(n + 2)\theta^n}.$$

We then find

$$\mathbf{E}\theta^2 = \mathbf{D}\theta + (\mathbf{E}\theta)^2 = \frac{\alpha a^2}{\alpha - 2}, \quad \mathbf{E}\theta^{-n} = \alpha a^\alpha \int_a^\infty \frac{d\theta}{\theta^{n+\alpha+1}} = \frac{\alpha}{(n + \alpha)a^n},$$

whence

$$\mathbf{E}(\mathbf{E}_\theta(\max (a, X_{(n)}))^2) = \frac{\alpha a^2(n + \alpha - 2)}{(n + \alpha)(\alpha - 2)}.$$

The risk of the Bayes estimate is

$$r(\delta^*) = \frac{\alpha a^2}{(\alpha - 2)(n + \alpha - 1)^2}.$$

We minimize $r(\delta^*) + cn$ in n and find the optimal number of observations

$$n^* = \left(\frac{2\alpha a^2}{c(\alpha - 2)}\right)^{1/3} - \alpha + 1.$$

5.23. We are seeking the value of $\mathbf{d}$ which minimizes the expectation

$$\begin{aligned}\mathbf{E}(L(\theta, \mathbf{d})|\mathbf{h}) &= \sum_{i=1}^{N} \mathbf{E}((\theta_i - d_i)^2|\mathbf{h}) \\ &= \sum_{i=1}^{N} \mathbf{E}((\theta_i - \mathbf{E}(\theta_i|\mathbf{h}))^2|\mathbf{h}) + \sum_{i=1}^{N} (\mathbf{E}(\theta_i|\mathbf{h}) - d_i)^2 \\ &= \sum_{i=1}^{N} \mathbf{D}(\theta_i|\mathbf{h}) + \sum_{i=1}^{N} (\mathbf{E}(\theta_i|\mathbf{h}) - d_i)^2.\end{aligned}$$

The minimum is attained for $d_i = \mathbf{E}(\theta_i|\mathbf{h})$, $i = 1, \ldots, N$, and is equal to $\sum_{i=1}^{N} \mathbf{D}(\theta_i|\mathbf{h})$. Here (see Problem 5.13 (6) and the hint)

$$\mathbf{E}(\theta_i|\mathbf{h}) = \frac{\alpha_i + h_i}{\alpha + n}, \; \mathbf{D}(\theta_i|\mathbf{h}) = \frac{(\alpha_i + h_i)(\alpha + n - \alpha_i - h_i)}{(\alpha + n)^2(\alpha + n + 1)}.$$

The first formula defines the Bayes estimate $\delta_i^*(\mathbf{h})$, and the second one allows us to calculate the respective risk $r(\delta^*) = \sum_{i=1}^{N} \mathbf{E}\mathbf{D}(\theta_i|\mathbf{h})$. We should also use the formulas (see Problem 1.52 and the hint)

$$\mathbf{E}h_i = \mathbf{E}(\mathbf{E}_\theta h_i) = \mathbf{E}n\theta_i = n\alpha_i/\alpha,$$

$$\mathbf{E}h_i(h_i - 1) = \mathbf{E}(\mathbf{E}_\theta h_i(h_i - 1)) = \mathbf{E}n(n - 1)\theta_i^2 = \frac{n(n - 1)\alpha_i(\alpha_i + 1)}{\alpha(\alpha + 1)}.$$

5.24. We use the notations of Problem 5.13 (7) and find by Problem 5.16 that the sought-for estimate is

$$\delta^*(\mathbf{x}) = \mathbf{E}(\theta|\mathbf{x}) = \mu_1 = \sigma_1^2\left(\frac{\mu}{\sigma^2} + \frac{x}{b^2}\right).$$

Here $\mathbf{D}(\theta|\mathbf{x}) = \sigma_1^2 \equiv \text{const}$, and the risk is $r(\delta^*) = (\sigma^{-2} + nb^{-2})^{-1}$. We minimize the total loss $r(\delta^*) + cn$ in n and find the optimal sample size

$$n^* = b\left(\frac{1}{\sqrt{c}} - \frac{b}{\sigma^2}\right).$$

5.25. (1) Let $d > d^*$. Then

$$|\theta - d| - |\theta - d^*| = \begin{cases} d^* - d & \text{for } \theta \geqslant d, \\ d + d^* - 2\theta & \text{for } d^* < \theta < d, \\ d - d^* & \text{for } \theta \leqslant d^*. \end{cases}$$

Since $d + d^* - 2\theta > d^* - d$ for $d^* < \theta < d$, we have

$$\begin{aligned} \mathbf{E}((|\theta - d| - |\theta - d^*|)|x) &\geqslant (d^* - d)\mathbf{P}(\theta \geqslant d|x) \\ &\quad + (d^* - d)\mathbf{P}(d^* < \theta < d|x) + (d - d^*)\mathbf{P}(\theta \leqslant d^*|x) \\ &= (d - d^*)(\mathbf{P}(\theta \leqslant d^*|x) - \mathbf{P}(\theta > d^*|x)). \end{aligned}$$

Since d^* is a median, the last difference is greater than or equal to zero. The case of $d < d^*$ is treated in a similar way. Thus, the decision d^* minimizes the average conditional (for $X = x$) loss and is a Bayes solution.

(2) By Problem 5.13 (7) the point $d^* = \mu_1$ is the median of the a posteriori distribution $\mathscr{L}(\theta|\mathbf{x})$. The Bayes estimate constructed from the sample $\mathbf{X} = (X_1, \ldots, X_n)$ has the form

$$\delta^*(\mathbf{X}) = \sigma_1^2\left(\frac{\mu}{\sigma^2} + \frac{n\overline{X}}{b^2}\right),$$

and its risk is

$$r(\delta^*) = \mathbf{E}|\theta - \delta^*(\mathbf{X})| = \mathbf{E}(\mathbf{E}(|\theta - \delta^*(\mathbf{X})||\mathbf{X})).$$

The conditional (for $\mathbf{X} = \mathbf{x}$) distribution of the random variable $\theta - \delta^*(\mathbf{X})$ is $\mathscr{N}(0, \sigma_1^2)$, and for $\mathscr{L}(Y) = \mathscr{N}(0, \sigma_1^2)$ we have $\mathbf{E}|Y| = \sqrt{\frac{2}{\pi}}\,\sigma_1$. Therefore,

$$r(\delta^*) = \sqrt{\frac{2}{\pi}}(\sigma^{-2} + nb^{-2})^{-1/2}.$$

If the price of one observation is $c > 0$, we minimize the total loss $r(\delta^*) + cn$ in n to find the optimal number of observations

$$n^* = b^2((\sqrt{2\pi}cb^2)^{2/3} - \sigma^{-2}).$$

TO CHAPTER 6

6.1. The representation

$$\mathbf{D}\overline{X} = \frac{1}{n^2} \sum_{k,s=1}^{n} \operatorname{cov}(X_k, X_s) = \frac{1}{n^2} \sum_{k,s=1}^{n} R_{k-s}$$

$$= \frac{1}{n}\left[R_0 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)R_k\right]$$

is true for the variance $\overline{X}$. Under the condition (6.1) it implies that $\mathbf{D}\overline{X} = O(1/n)$ as $n \to \infty$, i.e., $\overline{X}$ is a consistent estimate.

6.3. We can find the expression

$$\mathbf{E}\tilde{C}_k(n) = R_k - \frac{1}{n(n-k)} \sum_{t=1}^{n-k} \sum_{s=1}^{n} (R_{t-s} + R_{t+k-s}) + \frac{1}{n^2} \sum_{t,s=1}^{n} R_{t-s}$$

for the mean $\mathbf{E}\tilde{C}_k(n)$, where under condition (6.1) the term added to R_k is of the order $O(1/n)$ as $n \to \infty$.

6.4. Here $\mathbf{E}X_t = 0$ and

$$\operatorname{cov}(X_{k+t}, X_t) = \mathbf{E}(X_{k+t}, X_t)$$
$$= \sigma^2(\cos\lambda(k+t)\cos\lambda t + \sin\lambda(k+t)\sin\lambda t) = \sigma^2\cos\lambda\tilde{k}.$$

6.5. Here $\mathbf{E}X_t = m \sum_{j=0}^{r} \alpha_j$ and

$$\operatorname{cov}(X_{k+t}, X_t) = \sum_{i,j=0}^{r} \alpha_i\alpha_j \operatorname{cov}(\xi_{k+t-i}, \xi_{t-j})$$

$$= \begin{cases} \sigma^2 \sum_{j=0}^{r-|k|} \alpha_j\alpha_{j+|k|} & \text{for } |k| \leqslant r, \\ 0 & \text{for } |k| > r \end{cases}$$

because $\operatorname{cov}(\xi_{k+t-i}, \xi_{t-j}) = \sigma^2$ for $i - j = k$, while it is zero for $i - j \neq k$.

6.8. The coefficients of the optimal predictor are found from $\beta^*_{in} = \sum_{j=-n}^{0} R^{ij}R_{j-1}$, where $|R^{ij}| = |R_{i-j}|^{-1}$, $(i, j = 0, -1, \ldots, -n)$. For $\sigma^2(n)$ the representation $\sigma^2(n) = |\mathbf{R}(n+2)|/|\mathbf{R}(n+1)|$ is true (see [7, p. 263])

with the matrix

$$\mathbf{R}(n+1) = \begin{Vmatrix} R_0 R_1 \ldots\ldots R_n \\ R_1 R_0 \ldots\ldots R_{n-1} \\ \ldots\ldots\ldots\ldots\ldots \\ R_n R_{n-1} \ldots R_0 \end{Vmatrix}.$$

6.9. Since $\|p_{ij}(t)\| = \|p_{ij}(1)\|^t$, it is sufficient to verify that $\|p_{ij}(t+1)\| = \|p_{ij}(t)\| \|p_{ij}(1)\|$. The matrix $\|p_{ij}(1)\|$ is twice stochastic and, therefore, the stationary distribution is uniform.

6.10. Let $U, U_0, U_1, \ldots$ be a sequence of independent random variables uniformly distributed on the segment $[0, 1]$. We find the random variables

$$\zeta_0(U) = \begin{cases} 1 & \text{for} \quad U < 1/2, \quad \zeta^{(1)}(U) = \begin{cases} 1 & \text{for} \quad U < 1 - \alpha, \\ 2 & \text{for} \quad U \geqslant 1 - \alpha; \end{cases} \\ 2 & \text{for} \quad U \geqslant 1/2, \quad \zeta^{(2)}(U) = \begin{cases} 1 & \text{for} \quad U < \alpha, \\ 2 & \text{for} \quad U \geqslant \alpha. \end{cases} \end{cases}$$

Then the realization of the Markov chain is defined by

$$\nu_0 = \zeta_0(U_0), \; \nu_t = \zeta^{(\nu_{t-1})}(U_t), \; t \geqslant 1.$$

6.11. We have $\mathbf{E}\eta_t = 0$ and

$$R_k = \mathbf{E}\eta_{k+t}\eta_t = \mathbf{P}(\eta_{k+t} = \eta_t) - \mathbf{P}(\eta_{k+t} \neq \eta_t)$$
$$= \frac{1}{2}(p_{11}(k) + p_{22}(k) - p_{12}(k) - p_{21}(k)) = (1 - 2\alpha)^k.$$

6.13. Here $\mathbf{E}\eta_t = \mathbf{E}(\mathbf{E}(\xi_{\nu_t}(t) | \nu_t)) = 0$ because $\mathbf{E}\xi_i(t) = 0$. Then (see Problem 6.9) we have

$$\mathbf{E}\eta_{k+t}\eta_t = \mathbf{E}(\mathbf{E}(\xi_{\nu_{k+t}}(k+t)\xi_{\nu_t}(t) | \nu_{k+t}, \nu_t))$$
$$= \mathbf{P}(\nu_{k+t} = \nu_t = 1)R_k^{(1)} + \mathbf{P}(\nu_{k+t} = \nu_t = 2)R_k^{(2)}$$
$$= \frac{1}{4}(1 + (1 - 2\alpha)^k)(R_k^{(1)} + R_k^{(2)}).$$

Appendix

1. Normal Distribution

$$\text{Quantiles } p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_p} e^{-x^2/2}\,dx$$

p	u_p	p	u_p	p	u_p
0.50	0.000	0.68	0.468	0.86	1.080
0.51	0.025	0.69	0.496	0.87	1.126
0.52	0.050	0.70	0.524	0.88	1.175
0.53	0.075	0.71	0.553	0.89	1.227
0.54	0.100	0.72	0.583	0.90	1.282
0.55	0.126	0.73	0.613	0.91	1.341
0.56	0.151	0.74	0.643	0.92	1.405
0.57	0.176	0.75	0.674	0.93	1.476
0.58	0.202	0.76	0.706	0.94	1.555
0.59	0.228	0.77	0.739	0.95	1.645
0.60	0.253	0.78	0.772	0.96	1.751
0.61	0.279	0.79	0.806	0.97	1.881
0.62	0.305	0.80	0.842	0.98	2.054
0.63	0.332	0.81	0.878	0.99	2.326
0.64	0.358	0.82	0.915	0.999	3.090
0.65	0.385	0.83	0.954	0.9999	3.720
0.66	0.412	0.84	0.994	0.99999	4.265
0.67	0.440	0.85	1.036		

2. Poisson's Distribution

$$\text{Values of the function } \sum_{k=x}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$$

x \ λ	0.1	0.2	0.3	0.4	0.5	0.6
0	1.000	1.000	1.000	1.000	1.000	1.000
1	0.095	0.181	0.259	0.330	0.394	0.451
2	005	018	037	062	090	122
3		001	003	008	014	023
4				001	002	003

2. Poisson's Distribution (*cont.*)

x \ λ	0.7	0.8	0.9	1.0	2.0	3.0
0	1.000	1.000	1.000	1.000	1.000	1.000
1	0.503	0.551	0.593	0.632	0.865	0.950
2	156	191	228	264	594	801
3	034	047	063	080	323	577
4	006	009	014	019	143	353
5	001	001	002	004	053	185
6				001	018	084
7					005	034
8					001	012

x \ λ	4.0	5.0	6.0	7.0	8.0	9.0
0	1.000	1.000	1.000	1.000	1.000	1.000
1	0.982	0.993	0.998	0.999	1.000	1.000
2	908	960	983	924	0.997	0.999
3	762	875	938	970	986	994
4	567	735	849	918	958	979
5	371	560	715	827	900	945
6	215	384	554	699	809	884
7	111	238	394	550	687	793
8	051	138	256	401	547	676
9	021	068	153	271	408	544
10	008	032	084	170	283	413
11	003	014	043	099	184	294
12	001	005	020	053	112	197
13		002	008	027	068	124
14		001	004	013	034	074
15			001	006	017	042
16			001	002	008	022
17				001	004	011
18					002	005
19					001	002
20						001

3. Binomial Distribution

95%-confidence limits (θ_1, θ_2) for parameter θ:

$$P_\theta(\theta_1 < \theta < \theta_2) = 0.95$$

k \ $n-k$	0	1	2	3	4	5	6	7	8	9	10	11	12
0	—	0.98	0.84	0.71	0.60	0.52	0.46	0.41	0.37	0.34	0.31	0.29	0.27
	—	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	1.00	0.99	0.91	0.81	0.72	0.64	0.58	0.53	0.48	0.45	0.41	0.39	0.36
	0.03	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	1.00	0.99	0.93	0.85	0.78	0.71	0.65	0.60	0.56	0.52	0.48	0.45	0.43
	0.16	0.09	0.07	0.05	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02
3	1.00	0.99	0.95	0.88	0.82	0.76	0.70	0.65	0.61	0.57	0.54	0.51	0.48
	0.29	0.19	0.15	0.12	0.10	0.09	0.08	0.07	0.06	0.06	0.05	0.05	0.04
4	1.00	1.00	0.96	0.90	0.84	0.79	0.74	0.69	0.65	0.61	0.58	0.55	0.52
	0.40	0.29	0.22	0.18	0.16	0.14	0.12	0.11	0.10	0.09	0.08	0.08	0.07
5	1.00	1.00	0.96	0.92	0.86	0.81	0.77	0.72	0.68	0.65	0.62	0.59	0.56
	0.48	0.36	0.29	0.25	0.21	0.19	0.17	0.15	0.14	0.13	0.12	0.11	0.10
6	1.00	1.00	0.97	0.93	0.88	0.83	0.79	0.75	0.71	0.68	0.65	0.52	0.59
	0.54	0.42	0.35	0.30	0.26	0.23	0.21	0.19	0.18	0.16	0.15	0.14	0.13

3. Binomial Distribution (*cont.*)

k \ $n-k$	0	1	2	3	4	5	6	7	8	9	10	11	12
7	1.00	1.00	0.97	0.93	0.89	0.85	0.81	0.77	0.73	0.70	0.67	0.64	0.62
	0.59	0.47	0.40	0.35	0.31	0.28	0.25	0.23	0.21	0.20	0.18	0.17	0.16
8	1.00	1.00	0.98	0.94	0.90	0.86	0.82	0.79	0.75	0.72	0.69	0.67	0.64
	0.63	0.52	0.44	0.39	0.35	0.32	0.29	0.27	0.25	0.23	0.22	0.20	0.19
9	1.00	1.00	0.98	0.95	0.91	0.87	0.84	0.80	0.77	0.74	0.71	0.69	0.66
	0.66	0.56	0.48	0.43	0.39	0.35	0.32	0.30	0.28	0.26	0.24	0.23	0.22
10	1.00	1.00	0.98	0.95	0.92	0.88	0.85	0.82	0.79	0.76	0.73	0.70	0.68
	0.69	0.59	0.52	0.46	0.42	0.38	0.35	0.33	0.31	0.29	0.27	0.26	0.24
11	1.00	1.00	0.98	0.95	0.92	0.89	0.86	0.83	0.80	0.77	0.74	0.72	0.69
	0.72	0.62	0.57	0.49	0.45	0.41	0.38	0.36	0.34	0.32	0.30	0.28	0.27
12	1.00	1.00	0.98	0.96	0.93	0.90	0.87	0.84	0.81	0.78	0.76	0.73	0.71
	0.74	0.64	0.57	0.52	0.48	0.44	0.41	0.38	0.36	0.34	0.32	0.31	0.29

Note. The values of θ_2 are in the first rows, the values of θ_1 are in the second rows.

4. $\chi^2(n)$-Distribution

$$\text{Quantiles } p = \int_0^{\chi^2_{p,n}} k_n(x)\,dx = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^{\chi^2_{p,n}} x^{n/2-1} e^{-x/2}\,dx$$

n \ p	0.1	0.3	0.5	0.7	0.9	0.95	0.990	0.999
1	0.016	0.148	0.455	1.07	2.71	3.84	6.63	10.8
2	0.211	0.713	1.39	2.41	4.61	5.99	9.21	13.8
3	0.584	1.42	2.37	3.67	6.25	7.82	11.3	16.3
4	1.06	2.20	3.36	4.88	7.78	9.49	13.3	18.5
5	1.61	3.00	4.35	6.06	9.24	11.1	15.1	20.5
6	2.20	3.83	5.35	7.23	10.6	12.6	16.8	22.5
7	2.83	4.67	6.35	8.38	12.0	14.1	18.5	24.3
8	3.49	5.53	7.34	9.52	13.4	15.5	20.1	26.1
9	4.17	6.39	8.34	10.7	14.7	16.9	21.7	27.9
10	4.87	7.27	9.34	11.8	16.0	18.3	23.2	29.6
11	5.58	8.15	10.3	12.9	17.3	19.7	24.7	31.3
12	6.30	9.03	11.3	14.0	18.5	21.0	26.2	32.9
13	7.04	9.93	12.3	15.1	19.8	22.4	27.7	34.5
14	7.79	10.08	13.3	16.2	21.1	23.7	29.1	36.1
15	8.55	11.7	14.3	17.3	22.3	25.0	30.6	37.7
16	9.31	12.6	15.3	18.4	23.5	26.3	32.0	39.3
17	10.09	13.5	16.3	19.5	24.8	27.6	33.4	40.8
18	10.9	14.4	17.3	20.6	26.0	28.9	34.8	42.3
19	11.7	15.4	18.3	21.7	27.2	30.1	36.2	43.8
20	12.4	16.3	19.3	22.8	28.4	31.4	37.6	45.3
21	13.2	17.2	20.3	23.9	29.6	32.7	38.9	46.8
22	14.0	18.1	21.3	24.9	30.8	33.9	40.3	48.3
23	14.8	19.0	22.3	26.0	32.0	35.2	41.6	49.7
24	15.7	19.9	23.3	27.1	33.2	36.4	43.0	51.2
25	16.5	20.9	24.3	28.2	34.3	37.7	44.3	52.6
26	17.3	21.8	25.3	29.2	35.6	38.9	45.6	54.1
27	18.1	22.7	26.3	30.3	36.7	40.1	47.0	55.5
28	18.9	23.6	27.3	31.4	37.9	41.3	48.3	56.9
29	19.8	24.6	28.3	32.5	39.1	42.6	49.6	58.3
30	20.6	25.5	29.3	33.5	40.3	43.8	50.9	59.7

5. Student's Distribution $S(n)$

Values of the function $t'_{\gamma,n}$

$$\frac{1+\gamma}{2} = \int_{-\infty}^{t'_{\gamma,n}} s_n(x)\,dx = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \int_{-\infty}^{t'_{\gamma,n}} \left(1+\frac{x^2}{n}\right)^{-(n+1)/2} dx$$

n \ γ	0.9	0.95	0.98	0.99
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
12	1.782	2.179	2.681	3.055
14	1.761	2.145	2.625	2.977
16	1.746	2.120	2.584	2.921
18	1.734	2.101	2.552	2.878
20	1.725	2.086	2.528	2.845
22	1.717	2.074	2.508	2.819
24	1.711	2.064	2.492	2.797
26	1.706	2.056	2.479	2.779
28	1.701	2.048	2.467	2.763
30	1.697	2.042	2.457	2.750
∞	1.645	1.960	2.326	2.576

6. Snedecor's Distribution $S(n_1, n_2)$

Values of the function F_{p,n_1,n_2} for $p = 0.95$ and $p = 0.99$

$$p = \int_0^{F_{p,n_1,n_2}} f_{n_1,n_2}(x)\,dx$$

$$= \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \int_0^{F_{p,n_1,n_2}} x^{\frac{n_1}{2}-1}\left(1+\frac{n_1 x}{n_2}\right)^{-\frac{n_1+n_2}{2}} dx$$

The left limits of confidence intervals are found from the condition $F_{1-p,n_1,n_2} = F^{-1}_{p,n_2,n_1}$

n \ m	1	2	3	4	6	8	10	12	20	50	100
1	161 4052	200 4999	216 5403	225 5625	234 5859	239 5981	242 6056	244 6106	248 6208	252 6302	253 6334
2	18.51 98.50	19.00 99.01	19.16 99.17	19.25 99.25	19.33 99.33	19.37 99.36	19.39 99.40	19.41 99.42	19.44 99.45	19.47 99.48	19.49 99.49

6. Snedecor's Distribution $S(n_1, n_2)$ (*cont.*)

n \ m	1	2	3	4	6	8	10	12	20	50	100
3	10.13 34.12	9.55 30.81	9.28 29.46	9.12 28.71	8.94 27.91	8.84 27.49	8.78 27.23	8.74 27.05	8.66 26.69	8.58 26.35	8.56 26.23
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.16 15.21	6.04 14.80	5.96 14.54	5.91 14.37	5.80 14.02	5.70 13.69	5.66 13.57
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	4.95 10.67	4.82 10.27	4.74 10.05	4.68 9.89	4.56 9.55	4.44 9.24	4.40 9.13
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.28 8.47	4.15 8.10	4.06 7.87	4.00 7.72	3.87 7.39	3.75 7.09	3.71 6.99
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.58 6.37	3.44 6.03	3.34 5.82	3.28 5.67	3.15 5.36	3.03 5.06	2.98 4.96
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.22 5.39	3.07 5.06	2.97 4.85	2.91 4.71	2.77 4.41	2.64 4.12	2.59 4.01
12	4.75 9.33	3.88 6.93	3.49 5.95	4.26 5.41	3.00 4.82	2.85 4.50	2.76 4.30	2.69 4.16	2.54 3.86	2.40 3.56	2.35 3.46
20	4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.60 3.87	2.45 3.56	2.35 3.37	2.28 3.23	2.12 2.94	1.96 2.63	1.90 2.53

30	4.17	3.32	2.92	2.69	2.42	2.27	2.16	2.09	1.93	1.76	1.69
	7.56	5.39	4.51	4.02	3.47	3.17	2.98	2.84	2.55	2.24	2.13
50	4.03	3.18	2.79	2.56	2.29	2.13	2.02	1.95	1.78	1.60	1.52
	7.17	5.06	4.20	3.72	3.18	2.88	2.70	2.56	2.26	1.94	1.82
100	3.94	3.09	2.70	2.46	2.19	2.03	1.92	1.85	1.68	1.48	1.39
	6.90	4.82	3.98	3.51	2.99	2.69	2.51	2.36	2.06	1.73	1.59
200	3.89	3.04	2.65	2.41	2.14	1.98	1.87	1.80	1.62	1.42	1.32
	6.76	4.71	3.88	3.41	2.90	2.60	2.41	2.28	1.97	1.62	1.48
1000	3.85	3.00	2.61	2.38	2.10	1.95	1.84	1.76	1.58	1.36	1.26
	6.66	4.62	3.80	3.34	2.82	2.53	2.34	2.20	1.89	1.54	1.38

Notes. (1) The values of $F_{0.95,n_1,n_2}$ are in the first rows, the values of $F_{0.99,n_1,n_2}$ are in the second rows. (2) Here n_1 are the degrees of freedom for the greater variance, n_2 are the degrees of freedom for the smaller variance.

7. Kolmogorov's Test

Values of the function λ_p: $p = \mathbf{P}(D_n = \sup_x |F_n(x) - F(x)| > \lambda_p)$

$n \backslash p$	0.10	0.05	0.01	$n \backslash p$	0.10	0.05	0.01
1	0.950	0.975	0.995	19	0.271	0.301	0.361
2	776	842	929	20	265	294	352
3	636	708	829	25	238	264	317
4	565	624	734	30	218	242	290
5	509	563	669	35	202	224	269
6	468	519	617	40	189	210	252
7	436	483	576	45	179	198	238
8	410	454	542	50	170	188	226
9	387	430	513	55	162	180	216
10	369	409	489	60	155	172	207
11	352	391	468	65	149	166	199
12	338	375	449	70	144	160	192
13	325	361	432	75	139	154	185
14	314	349	418	80	135	150	179
15	304	338	404	85	131	145	174
16	295	327	392	90	127	141	169
17	286	318	381	95	124	137	165
18	279	309	371	100	121	134	161

8. Smirnov's Test

The probabilities $\mathbf{P}(D_{nn} \leqslant k/n)$, where $D_{nn} = \sup_x |F_{1n}(x) - F_{2n}(x)|$

$k \backslash n$	1	2	3	4	5	6	7	8	9	10	11
1	1.000										
2	0.667	1.000									
3	400	0.900	1.000								
4	229	771	0.971	1.000							
5	127	643	921	0.992	1.000						
6	069	526	857	974	0.998	1.000					
7	037	425	788	947	992	0.999	1.000				
8	020	340	717	913	981	998	1.000	1.000			
9	011	270	648	874	966	993	0.999	1.000	1.000		
10	006	213	582	832	948	988	998	1.000	1.000	1.000	
11	003	167	521	789	925	979	996	0.999	1.000	1.000	1.000
12	002	131	464	744	900	969	992	999	1.000	1.000	1.000
13	001	102	412	700	874	956	987	997	1.000	1.000	1.000
14	000	079	365	657	845	941	981	995	0.999	1.000	1.000
15	000	062	322	614	816	925	974	992	998	1.000	1.000
16	000	048	284	574	785	907	965	989	997	1.000	1.000
17	000	037	249	535	755	888	955	984	995	0.999	0.999
18	000	028	219	497	725	868	944	979	993	998	999
19	000	022	192	462	694	847	932	973	991	997	999
20	000	017	168	429	664	825	919	966	988	996	999
21	000	013	147	397	635	804	905	959	984	995	998
22	000	010	128	368	606	782	891	951	980	993	998
23	000	008	112	340	578	759	876	942	975	991	997
24	000	006	098	314	551	737	860	932	970	988	996

9. Random Numbers Uniformly Distributed on [0, 1]

0.5916	3406	6079	4101	5314
6562	7463	8203	1643	5825
3127	1413	9711	6253	4135
0690	0120	3993	3136	3821
3617	6700	5940	9629	1094
7128	6396	6787	3147	2625
6635	5477	9121	4513	6213
9162	3901	7480	6319	2645
9313	5889	0399	2226	7919
8216	8851	4184	0471	9664
9470	2099	1992	0836	5050
3361	6387	3374	4963	1255
5303	5501	4237	5307	8954
1039	9430	6838	4188	2383
9031	4215	1197	8764	8382
9481	4474	8315	1752	8546
8922	6145	5759	5489	1479
5725	6542	2141	7449	1653
4398	7198	5643	3687	2311
3652	5889	8865	2378	2198
9612	0448	9632	3741	4776
6836	0101	8861	2786	5132
4601	8247	6883	2196	6570
9154	7397	3584	2139	1019
2212	8036	6484	9953	8382
7158	2036	5270	7441	4387
9192	9019	7880	4728	0115
3072	2267	6512	5673	2943
2380	4955	7803	1907	5803
3290	8562	2558	5986	1904
4448	1790	1932	0833	7005
7042	4161	9279	4049	1693
5978	5412	2154	9202	7586
7147	7403	5033	8549	6005
4386	9362	6122	0193	1987

10. Random Numbers Normally Distributed as $\mathcal{N}(0, 1)$

−0.486	0.856	−0.491	−1.983	−1.787	−0.261
−0.256	−0.212	0.219	0.779	−0.105	−0.357
0.065	0.415	−0.169	0.313	−1.339	1.827
1.147	−0.121	1.096	0.181	1.041	0.535
−0.199	−0.246	1.239	−2.574	0.279	−2.056
1.237	1.046	−0.508	−1.630	−0.146	−0.392
−1.384	0.360	−0.992	−0.116	−1.698	−2.832
−0.959	0.424	0.969	−1.141	−1.041	0.362
0.731	1.377	0.983	−1.330	1.620	−1.040
0.717	−0.873	−1.096	−1.396	1.047	0.089
−1.805	−2.008	−1.633	0.542	0.250	−0.166
−1.186	1.180	1.114	0.882	1.265	−0.202
0.658	−1.141	1.151	−1.210	−0.927	0.425
−0.439	0.358	−1.939	0.891	−0.227	0.602
−1.399	−0.230	0.385	−0.649	−0.577	0.237
0.032	0.079	0.199	0.208	−1.083	−0.219
0.151	−0.376	0.159	0.272	−0.313	0.084
0.290	−0.902	2.273	0.606	0.606	−0.747
0.873	−0.437	0.041	−0.307	0.121	0.790
−0.289	0.513	−1.132	−2.098	0.921	0.145
−0.291	1.221	1.119	0.004	0.768	0.079
−2.828	−0.439	−0.792	−1.275	0.375	−1.656
0.247	1.291	0.063	−1.793	−0.513	−0.344
−0.584	0.541	0.484	−0.986	0.292	−0.521
0.446	−1.661	1.045	−1.363	1.026	2.990
0.034	−2.127	0.665	0.084	−0.880	−1.473
0.234	−0.656	0.340	−0.086	−0.158	−0.851
−0.736	1.041	0.008	0.427	−0.831	0.210
−1.206	−0.899	0.110	−0.528	−0.813	1.266
−0.491	−1.114	1.297	−1.433	−1.345	−0.574
−1.334	1.278	−0.568	−0.109	−0.515	−0.566
−0.287	−0.144	−0.254	0.574	−0.451	−1.181
0.161	−0.886	−0.921	−0.509	1.410	−0.518
−1.346	0.193	−1.202	0.394	−1.045	0.843
1.250	−0.199	−0.288	1.810	1.378	0.584

10. Random Numbers Normally Distributed as $\mathcal{N}(0, 1)$ (*cont.*)

2.923	0.500	0.630	−0.537	0.782	0.060
−1.190	−0.318	0.375	−1.941	0.247	−0.491
0.192	−0.432	−1.420	0.489	−1.711	−1.186
0.942	1.045	−0.151	−0.243	−0.430	−0.762
1.216	0.733	−0.309	0.531	0.416	−1.541
0.499	−0.431	1.705	1.164	0.424	−0.444
0.665	−0.135	−0.145	−0.498	0.593	0.658
0.754	−0.732	−0.066	1.006	0.862	−0.885
0.298	1.049	1.810	2.885	0.235	−0.628
1.456	2.040	−0.124	0.196	−0.853	0.402
0.593	0.993	−0.106	0.116	0.484	−1.272
−1.127	−1.407	−1.579	−1.616	1.458	1.262
−0.142	−0.504	0.532	1.381	0.022	−0.281
−0.023	−0.463	−0.899	−0.394	−0.538	1.707
0.777	0.833	0.410	−0.349	−1.094	0.580
0.241	−0.957	−1.885	0.371	−2.830	−0.238
0.022	0.525	−0.255	−0.702	0.953	−0.869
−0.853	−1.865	−0.423	−0.973	−1.016	−1.726
−0.501	−0.273	0.857	−0.465	−1.691	0.417
0.439	−0.035	−0.260	0.120	−9.558	0.056

List of Distributions

1. Normal $\mathcal{N}(\mu, \sigma^2)$	**1**: 9, 10, 11, 27, 29, 32, 39, 41, 43, 46, 47, 56, 57, 58, 60, 61. **2**: 13, 14, 15, 16, 17, 18, 19, 20, 43, 48, 50, 52, 53, 64, 65, 66, 67, 68, 70, 71, 72, 84, 85, 86, 87, 88, 89, 94, 106, 109, 110, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 131, 142, 143, 145, 146, 148. **3**: 35, 45, 47, 48, 49, 50, 58, 59, 60, 61, 65, 66, 67, 73, 74, 79, 80, 84, 87. **4**: 8, 9, 10, 11, 13, 14, 15, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 34, 36. **5**: 7, 8, 13, 24, 25.
2. Normal multivariate $\mathcal{N}(\mu, \Sigma)$	**1**: 28, 40, 53, 59, 62, 63. **2**: 39, 44, 90, 91, 92, 93, 132, 150. **3**: 52, 72, 88. **5**: 11, 12.
3. Binomial $Bi(n, p)$	**1**: 1, 2, 12, 13, 14, 15, 16, 17, 18, 39, 52, 54. **2**: 5, 6, 7, 8, 43, 48, 57, 84, 109, 110, 133, 134, 135, 136, 148. **3**: 1, 2, 5, 17, 18, 39, 40, 46, 53, 63, 75, 77. **5**: 1, 2, 3, 4, 13, 14, 15, 16.
4. Polynomial $M(n; p_1, \ldots, p_N)$	**1**: 3, 19, 20, 26, 39, 52, 53, 54. **2**: 29, 38, 45, 63, 144. **3**: 3, 4, 7, 8, 9, 12, 13, 20, 24, 34, 37, 38. **5**: 13, 23.
5. Poisson's $\Pi(\lambda)$	**1**: 39, 54, 55, 60, 64. **2**: 9, 10, 31, 43, 48, 58, 59, 61, 84, 97, 108, 109, 137, 138, 139. **3**: 14, 15, 16, 41, 54, 64, 76, 78. **5**: 6, 11, 13, 18, 19.
6. Binomial negative $\overline{Bi}(r, p)$	**1**: 39, 55. **2**: 11, 12, 43, 48, 62, 84, 96, 140. **3**: 42, 55. **5**: 5, 13, 17.
7. Gamma $\Gamma(a, \lambda)$	**1**: 6, 7, 8, 21, 34, 39, 42, 44, 51, 55. **2**: 21, 22, 30, 43, 48, 51, 73, 74, 75, 84, 104, 109, 127, 128, 130, 141. **3**: 10, 11, 43, 56, 62, 68, 69. **5**: 13, 18, 20.

8. Uniform $R(a, b)$	**1**: 5, 19, 35, 36, 43, 46. **2**: 23, 24, 25, 32, 79, 80, 81, 100, 101, 107, 110, 129, 148. **3**: 25, 36, 45, 70. **4**: 7, 12, 33, 35. **5**: 13, 21, 22.
9. Weibull's $W(a, \alpha, b)$	**1**: 37. **2**: 26, 76, 77, 102, 103, 107, 130. **3**: 71.
10. Cauchy's $C(a)$	**1**: 46. **2**: 28, 43, 99. **3**: 44.
11. Hypergeometric $H(r, N, n)$	**2**: 33, 113. **3**: 57.
12. Chi-square $\chi^2(n)$	**1**: 40, 45, 47, 51, 57, 59, 147. **3**: 19.
13. Beta $B(a, b)$	**1**: 44, 47, 48, 49. **2**: 48. **5**: 13, 17.
14. Student's $S(n)$	**1**: 47, 50, 59.
15. Snedecor's $S(n_1, n_2)$	**1**: 48, 49, 50, 51.
16. Logistic	**2**: 27, 49.
17. Power series	**2**: 60, 96, 140.
18. Pareto's $\Pi(a, \alpha)$	**5**: 13, 21.
19. Dirichlet's $D(\alpha)$	**5**: 13, 23.
20. Laplace	**2**: 105.
21. Kapteyn's	**2**: 95.
22. Gaussian inverse	**2**: 54.
23. Finite population	**2**: 34, 35, 36, 37, 83, 111, 112.

Bibliography

1. Bickel P. J. and Doksum K. A. *Mathematical Statistics. Basic Ideas and Selected Topics*. Holden-Day, San Francisco, etc., 1977.

2. Chistyakov V. P. *A Course in Probability Theory*. Nauka, Moscow, 1987 (in Russian).

3. Cramér H. *Mathematical Methods of Statistics*. Princeton, 1946.

4. Ermakov S. M. and Mikhailov G. A. *Statistical Simulation*. Nauka, Moscow, 1982 (in Russian).

5. Feller W. *An Introduction to Probability Theory and Its Applications*. Vol. 1, Third Edition, J. Wiley & Sons, New York, 1968.

6. Ivchenko G. I., Glibochenko A. F., Ivanov V. A., and Medvedev Yu. I. *Statistical Analysis of Discrete Random Sequences*. Moscow Institute of Electronic Engineering, Moscow, 1984 (in Russian).

7. Ivchenko G. I. and Medvedev Yu. I. *Mathematical Statistics*. Mir, Moscow, 1990.

8. Kendall M. G. and Stuart A. *The Advanced Theory of Statistics*. Vol. 1. *Distribution Theory*. Third Edition, Griffin, London, 1969.

9. Knuth D. E. *The Art of Computer Programming*. Vol. 2, Second Edition, Eddison Wesley, Reading (Mass.), etc., 1981.

10. Lehman E. L. *Testing Statistical Hypotheses*. J. Wiley & Sons, New York, 1959.

To the Reader

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:

Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR